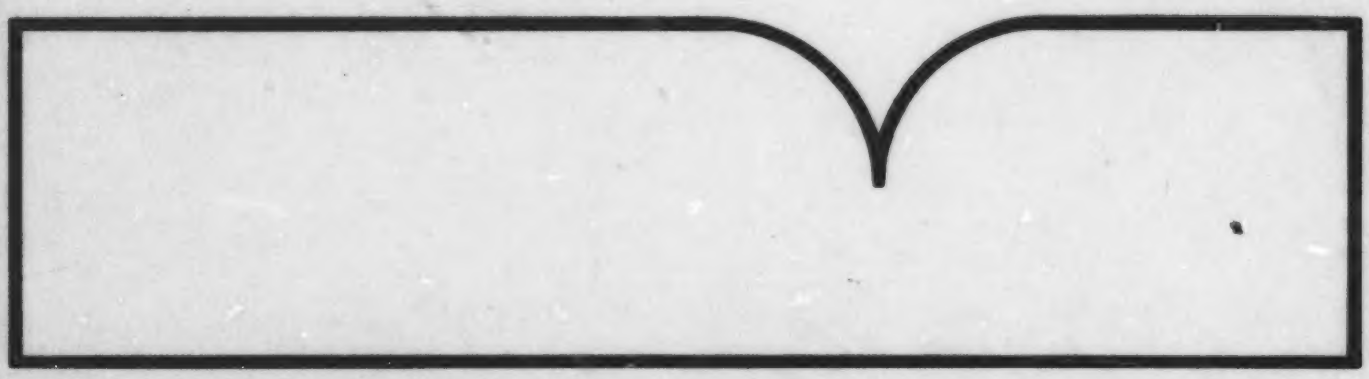
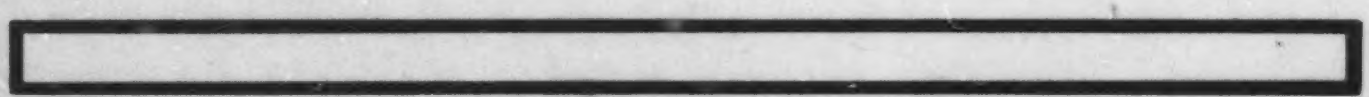
PB81-172660

Disfit: A Distribution Fitting System 1 Discrete Distributions, A User's Guide

(U.S.) Health Effects Research Lab.
Research Triangle Park, NC

Jan 81





U.S. Department of Commerce
National Technical Information Service
NTIS.

TECHNICAL REPORT DATA (Please read Instructions on the reverse before completing)		
1. REPORT NO. EPA-600/2-81-010	2.	3. RECIPIENT'S ACCESSION NO. PB81 172660
4. TITLE AND SUBTITLE Disfit: A Distribution Fitting System 1. Discrete Distribution		5. REPORT DATE January 1981 Issuing Date.
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Victor Hasselblad, Andy Stead, Helen Anderson		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Biometry Division Health Effects Research Laboratory U.S. Environmental Protection Agency Research Triangle Park, NC 27711		10. PROGRAM ELEMENT NO. 1AA817
		11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Office of Research and Development Health Effects Research Laboratory U.S. Environmental Protection Agency Research Triangle Park, NC 27711		13. TYPE OF REPORT AND PERIOD COVERED User's Guide
		14. SPONSORING AGENCY CODE EPA 600/11
15. SUPPLEMENTARY NOTES		

16. ABSTRACT

The DISFIT system is a series of programs and subroutines to fit distributions to data. This first volume describes the routines to fit discrete distributions. The distributions included are the binomial, truncated binomial, mixture of two binomials, beta binomial, Poisson, truncated Poisson, mixture of two Poissons, negative binomial, truncated negative binomial, and logarithmic. All parameters are estimated using maximum likelihood techniques. Any of the parameters may be specified instead of estimated. Variances of estimated asymptotic variances of the parameter estimates are also given. Some tests of hypotheses are possible using likelihood ratio tests. This guide contains the descriptions, calling sequences, documentation, and examples for each distribution. The program is written entirely in FORTRAN, and a listing of the program is in the appendix.

17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS (This Report) UNCLASSIFIED	21. NO. OF PAGES 85
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (9-73)

TYPING GUIDE SHEET
BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
DROPPED HEAD; BEGIN SECTIONS HERE

EPA 600/2-81-010
January 1981

PB81-172660

DISFIT: A DISTRIBUTION FITTING SYSTEM

1. DISCRETE DISTRIBUTIONS

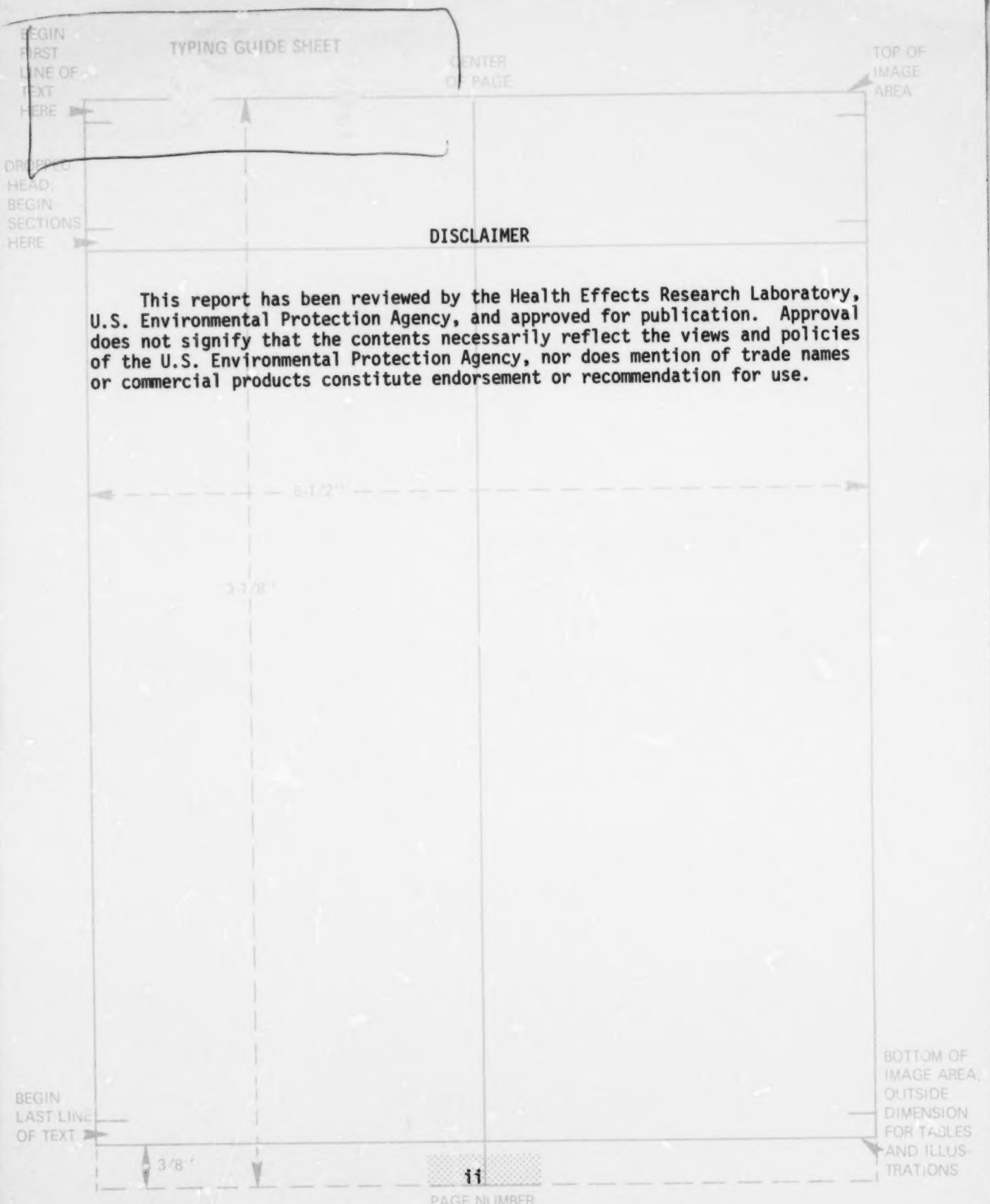
A User's Guide

Victor Hasselblad
Andrew G. Stead
Helen S. Anderson

BIOMETRY DIVISION
HEALTH EFFECTS RESEARCH LABORATORY
U. S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27711

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA 22161

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
6-1/2"
1-1/8"
3/8"
PAGE NUMBER
EPA-287 (Cin.)
(4-76)



DISCLAIMER

This report has been reviewed by the Health Effects Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

CONTENTS

1. Introduction 1
2. Methods of Estimation 3
3. Use of the Program 4
4. Constant Distribution. 5
5. Binomial Distribution. 8
6. Truncated Binomial Distribution. 11
7. Mixture of Two Binomial Distributions. 14
8. Beta-Binomial Distribution 17
9. Poisson Distribution 21
10. Truncated Poisson Distribution 24
11. Mixture of Two Poisson Distributions 27
12. Negative Binomial Distribution 30
13. Truncated Negative Binomial Distribution 34
14. Logarithmic Distribution 37
15. References 40
Appendix A. Index 43
Appendix B. Flow Chart of the Program 45
Appendix C. Error Messages. 46
Appendix D. Fortran Program Listing 47

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin.)
(4-76)

TYPING GUIDE SHEET
BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE
9-1/8"
BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA. OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"
PAGE NUMBER
EPA-287 (Cin.) (4-76)

ABSTRACT

The DISFIT system is a series of programs and subroutines to fit distributions to data. This first volume describes the routines to fit discrete distributions. The distributions included are the binomial, truncated binomial, mixture of two binomials, beta binomial, Poisson, truncated Poisson, mixture of two Poissons, negative binomial, truncated negative binomial, and logarithmic. All parameters are estimated using maximum likelihood techniques. Any of the parameters may be specified instead of estimated. Variances or estimated asymptotic variances of the parameter estimates are also given. Some tests of hypotheses are possible using likelihood ratio tests. This guide contains the descriptions, calling sequences, documentation, and examples for each distribution. The program is written entirely in FORTRAN, and a listing of the program is in the appendix.

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE
8-1/2"
9-1/8"
BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"
PAGE NUMBER
EPA-287 (Cin.)
(4-76)

ACKNOWLEDGMENT

The authors would like to express their appreciation to Dave Mage and Gene Lowrimore for their helpful suggestions. We are especially grateful to Barbara Crabtree for her typing and editing of this guide.

Page vi blank

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD, BEGIN SECTIONS HERE

1. INTRODUCTION

Although distribution fitting methods abound in the statistical literature, very few of these methods are found in the major statistical packages. In particular, SPSS [27], BMD-P [10] and SAS [3] only give some overall tests for normality. There are at least three specialized distribution fitting packages. Bouver and Bargmann [6] wrote a system to fit the Pearson system of curves by the method of moments. The system of Alldredge and Bolstad [1] fits the normal, lognormal, exponential, gamma, and uniform distributions using maximum likelihood for some distributions, and moment estimators for others. Morgan, Nelson, and Caporal [25] have written a system that includes the normal, lognormal, Weibul and binomial. This system has the advantage that linear models can be fitted to the shape parameter, and maximum likelihood methods are used. This feature limits the kinds of distributions which are fitted, however. None of these systems include a wide variety of distributions commonly appearing in the literature.

The examples chosen for this distribution fitting package include some of the classic ones, as well as those pertinent to the environmental field. The analysis of both environmental monitoring and environmental health data require distribution fitting programs. For example, ambient air standards are often based on the distribution of the "second maximum". Health surveys often yield counts, and the distributional form of these counts may be important for hypothesis testing.

This first report will be restricted to discrete distributions. Much of the information on the distributions is found in "Discrete Distributions" by Johnson and Kotz [19]. This reference is extremely comprehensive, so only a small portion of its information is contained in this report. In a manner similar to Johnson and Kotz [19], each section of this report discusses a particular distribution. Included are definitions of the distribution, estimating equations, goodness-of-fit tests, and actual examples. In general, parameter estimates and their variances are given, but tests of hypotheses are not usually given. Tests of hypotheses often can be made from the likelihood functions provided. The appendix of the paper contains the FORTRAN routines to perform the estimation and distribution fitting. A "main" program is also included to read in the data, and to enable the user to fit several different distributions to the same data.

The routines are written so that each parameter may be specified or estimated. In some cases, specifying a parameter value may give a specific distribution. For example, if N is set to 1, the negative binomial distribution becomes the geometric distribution. There may be other cases

BEGIN LAST LINE OF TEXT HERE
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin.) (4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE
6-1/2"
8-1/8"
BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"
PAGE NUMBER
EPA-287 (Cin.)
(4-76)

where one of the parameters is known, and the other(s) need to be estimated.

We welcome suggestions and corrections to this report and its programs. Two additional reports are planned. They will contain continuous distributions and grouped continuous distributions respectively.

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

2. METHODS OF ESTIMATION

Estimates in this paper are calculated by maximum likelihood. Any estimates calculated by other means will be so indicated. There are, of course, several reasons for using maximum likelihood estimates. The estimates have several optimal properties including asymptotic consistency and efficiency (Kendall and Stuart [20]). The comparison of the likelihood function for different distributions can give some measure of goodness-of-fit. In some specific cases, it is possible to test the fit of one distribution versus another using likelihood ratio tests.

Maximum likelihood estimates also have certain problems. Some of these are caused by small sample sizes, although distribution fitting usually requires moderate to large sample sizes. The likelihood function for mixtures of continuous distributions can have singularities corresponding to each data point, as shown by Day [8]. The discrete distributions considered in this paper do not have this particular problem. Finally, the likelihood function may have more than one local maximum. This can occur even with discrete distributions, and we can only hope that the initial estimates lead to meaningful solutions in each case.

Distributions such as the gamma or beta, or functions of them such as the negative binomial, require iterative methods to obtain maximum likelihood estimates. The modified Gauss-Newton method is easy to compute, and has the advantage of being a second order method. A description of the method was given by Hartley [15]. In other cases the standard Newton-Raphson method was used.

Maximum likelihood estimation is especially useful for problems of "missing information". Missing information can take the form of censored data, truncated distributions, grouping, and mixing of distributions. All of these problems can be handled by using the "missing information principle" (MIP) of Orchard and Woodbury [28]. This method will converge under a wide variety of conditions, and the convergence is geometric as has been shown by Dempster, Laird, and Rubin [9]. We have also used this method in our routines.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

3. USE OF THE PROGRAM

The program to fit various discrete distributions was written entirely in FORTRAN 77 on a Univac 1100 computer. It was written so that conversion to other FORTRAN compilers should be relatively simple. The program consists of a main program to read in the data, and a series of subroutines - one for each different distribution. The exact calling sequence for the main program will depend on the job control language of the computer used.

The main program reads in the following cards:

Variable	Discription	Format
TITLE	80 character title	20A4
K	Value of largest count	I5
FMT	Variable format statement using integer mode only	20A4
NF(I)	K counts, NF(I),I=1,K+1, are read in, where NF(1) = no. of 0's, NF(2) = no. of 1's, NF(3) = no. of 2's,...,and NF(K+1) = no. of values = K.	FMT

These counts, NF(I) will be denoted by f_{i-1} in the formulas in the text, so that f_i denotes the number of occurrences of the value "i". For an explanation of a variable format statement, see any FORTRAN manual, or the SPSS [27] manual.

This information is followed by a card for each distribution to be fitted. The description of each card is given in the chapter corresponding to the distribution requested.

The following example shows the cards just described.

```
WELDON'S DICE DATA
   12
(11I5)
  185 1149 3265 5475 6114 5194 3067 1331  403  105   14    4    0
BINOMIAL
POISSON
```

A blank card following the distribution cards finishes the analyses of the data set. The program will then attempt to read in additional data sets. An end-of-file card will terminate the run.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

4. CONSTANT DISTRIBUTION

4.1 Definition

The "constant" distribution routine is really a summary routine. It computes the first four sample moments:

$$M_1 = \sum_{i=0}^{K} i\, f_i/n, \text{ where } n = \sum_{i=0}^{K} f_i, \text{ and} \qquad (4.1)$$

$$M_j = \sum_{i=0}^{K} (i-M_1)^j f_i/n, \; j = 2,3,4 \qquad (4.2)$$

The routine also prints the frequency and cumulative frequency distribution.

It is also possible to specify the expected frequencies through constants (thus the name of the routine), C_i's, $i = 0,1,\ldots,K$. These constants are normalized so that the sum of the C_i's = n. Thus the C_i's can be expected counts, probabilities, or just proportional numbers. If the C_i's are not specified, they are assumed to be equal. This particular case is commonly called the discrete uniform or discrete rectangular distribution.

4.2 Calling Sequence

The constant routine is called by

Col. 8	15	25	35	45	55
CONSTANT	<NP>	$<C_1>$	$<C_1>$	$<C_2>$	$<C_3>$

If no constants are specified, then each C_i is set to n/(K+1). If more than four constants are needed (K > 3), they are read in on additional cards using the format (8F10.0). NP is the number of parameters estimated from the data itself, and is often left blank.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

4.3 Tests for Goodness-of-Fit

A standard chi-square goodness-of-fit test is computed for the observed frequencies:

$$X^2 = \sum_{i=0}^{K} (f_i - C_i)^2 / C_i \qquad (4.3)$$

which has K-NP degrees of freedom. The cells are collapsed from the left and right until the end cells have an expected value of 5. The degrees of freedom will be adjusted for collapsing. The log-likelihood function is also computed:

$$L = \sum_{i=0}^{K} f_i \log(C_i/n) \qquad (4.4)$$

4.4 Example

De Winton and Bateson collected counts of Primula in a cross breeding experiment. The results are given by Fisher [13]. The expectations involving two independent Mendelian factors would produce counts in the ratio of 9:3:3:1. Figure 4.1 shows that the fit using this ratio is not good. If the first two counts came from a more viable plant, then the ratios for the first two cells still would be 3:1 as would be the ratio for the last two. The relative ratio of the first two cells to the last two cells could be estimated from the data, giving the expected counts in Figure 4.2. Since one parameter is being estimated from the data, this is indicated on the control card corresponding to Figure 4.2. The result is a satisfactory fit. A likelihood ratio test for the difference in fit can be computed as -2(-613.34 + 608.82) = 9.04. This is asymptotically distributed as a chi-square with one degree of freedom, which corresponds to a p-value of .0026.

Input for Primula data:

```
DE WINTON AND BATESON PRIMULA DATA
    3
(4I5)
  328  122   77   33
CONSTANT          9.          3.          3.          1.
CONSTANT          1       337.5       112.5        82.5        27.5
```

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

Outputs:

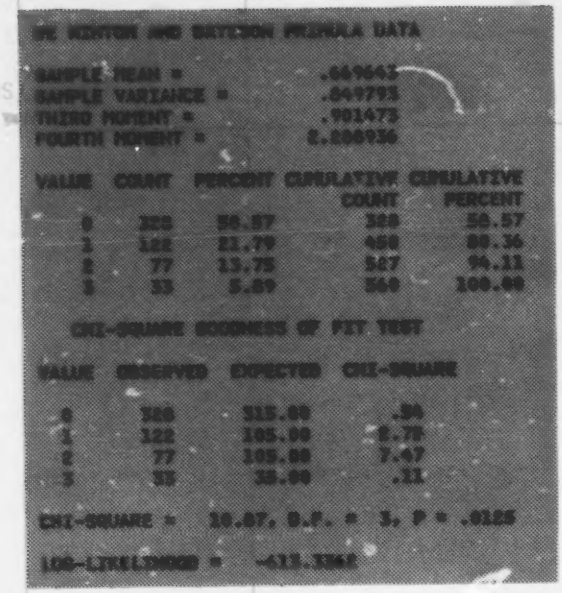

SAMPLE MEAN = .669643
SAMPLE VARIANCE = .849795
THIRD MOMENT = .981473
FOURTH MOMENT = 2.280936
VALUE COUNT PERCENT CUMULATIVE COUNT CUMULATIVE PERCENT
0 328 56.87 328 56.87
1 122 21.79 450 78.66
2 77 13.75 527 94.11
3 33 5.89 560 100.00
CHI-SQUARE GOODNESS OF FIT TEST
VALUE OBSERVED EXPECTED CHI-SQUARE
0 328 315.00 .90
1 122 105.00 2.76
2 77 105.00 7.47
3 33 35.00 .11
CHI-SQUARE = 10.87, D.F. = 3, P = .0125
LOG-LIKELIHOOD = -413.3368

Figure 4.1

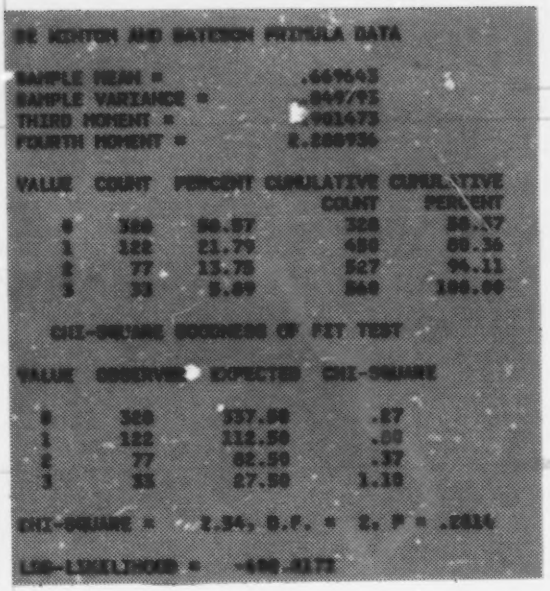

SAMPLE MEAN = .669643
SAMPLE VARIANCE = .849795
THIRD MOMENT = .981473
FOURTH MOMENT = 2.280936
VALUE COUNT PERCENT CUMULATIVE COUNT CUMULATIVE PERCENT
0 328 56.87 328 56.87
1 122 21.79 450 78.66
2 77 13.75 527 94.11
3 33 5.89 560 100.00
CHI-SQUARE GOODNESS OF FIT TEST
VALUE OBSERVED EXPECTED CHI-SQUARE
0 328 337.50 .87
1 122 112.50 .89
2 77 82.50 .37
3 33 27.50 1.10
CHI-SQUARE = 2.34, D.F. = 2, P = .2814

Figure 4.2

6-1/8"
BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

5. BINOMIAL DISTRIBUTION

5.1 History and Definition

The binomial distribution was derived by James Bernoulli in 1713. The variable, x, has the possible values of 0, 1, 2, ... , N, where the probability of x is given by

$$\Pr[x=j] = \binom{N}{j} p^j q^{N-j}, \tag{5.1}$$

subject to $q = 1 - p$ and $0 < p < 1$. The number N will be assumed to be constant and known. It is assumed that the distribution is sampled n times, so that the outcome i occurs f_i times. Thus the f_i's sum to n. In some data, the value N varies for each sample. We have not allowed for this possibility since it would require a very different data structure. The estimation equations for this and related distributions are easily generalized from the equations given.

5.2 Calling Sequence

The binomial distribution routine is called by

```
|Col.  8|   15|      25|
|BINOMIAL|  <N>|     <P>|
```

where N is an integer, and P is the value of p to be fitted. If N is not specified, it will be set to K, the value of the largest count. If p is not specified, it will be estimated from the data.

5.3 Maximum Likelihood Estimate

The maximum likelihood estimate for p is

$$\hat{p} = \sum_{i=0}^{K} i\, f_i/(nN) \tag{5.2}$$

The estimate is also the minimum variance unbiased estimate. The variance of $\hat{p}$ is estimated by

$$\mathrm{Var}(\hat{p}) = \hat{p}(1-\hat{p})/(nN) \tag{5.3}$$

BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
BEGIN LAST LINE OF TEXT
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

5.4 Test for Goodness-of-Fit

A standard goodness-of-fit test is computed for the observed frequencies

$$X^2 = \sum_{i=0}^{K}(f_i - E_i)^2/E_i, \text{ where } E_i = nPr[x=i] \qquad (5.4)$$

and the cells are collapsed from the left and right until the end cells each have an expected value of 5. The degrees of freedom will be the number of cells after collapsing minus two. If p is specified, then only one will be subtracted.

A second goodness-of-fit test is commonly known as the homogeneity test. It is a test of the variance of the sample, and is discussed by Potthoff and Whittinghill [30].

$$X^2 = (\sum_{i=0}^{K} f_i^2)/(\hat{p}(1-\hat{p})) - N\hat{p}/(1-\hat{p}). \qquad (5.5)$$

The statistic is distributed approximately as a χ^2 with n-1 degrees of freedom if p is estimated, or n degrees of freedom if p is specified.

5.5 Example

Fisher [13] also gives counts from a dice experiment of Weldon. The counts represent the number of 5's and 6's occuring in the toss of 12 dice. The distribution of this number should be given by a binomial distribution where $p = 1/3$. In fact, using a p of 1/3 gives a very poor fit (Figure 5.1). If instead p is estimated from the data, the estimated p of .3377 gives a very adequate fit. A likelihood ratio test of $p = 1/3$ gives an asymptotic chi-square of -2(-50255.16 + 50241.65) = 27.02, which corresponds to a p-value of less than .0001.

Input for Weldon's dice data:

```
WELDON'S DICE DATA
   12
(13I5)
  185 1149 3265 5475 6114 5194 3067 1331  403  105   14    4    0
BINOMIAL          0.333333
BINOMIAL
```

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

Outputs:

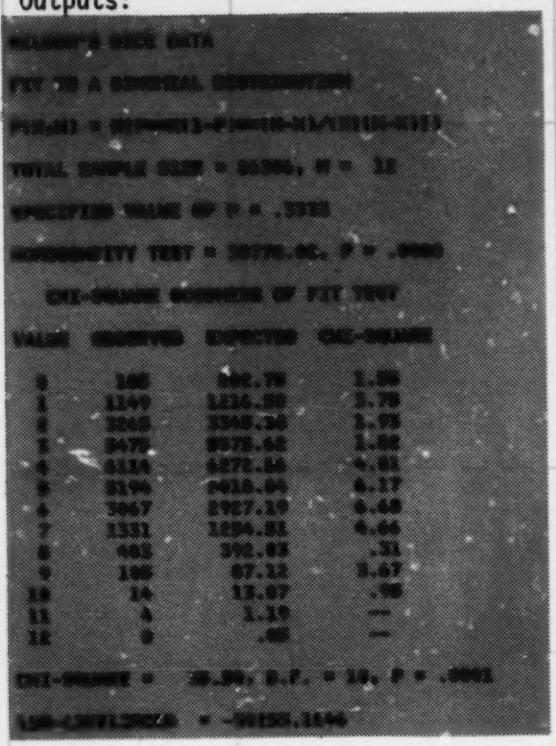


Figure 5.1

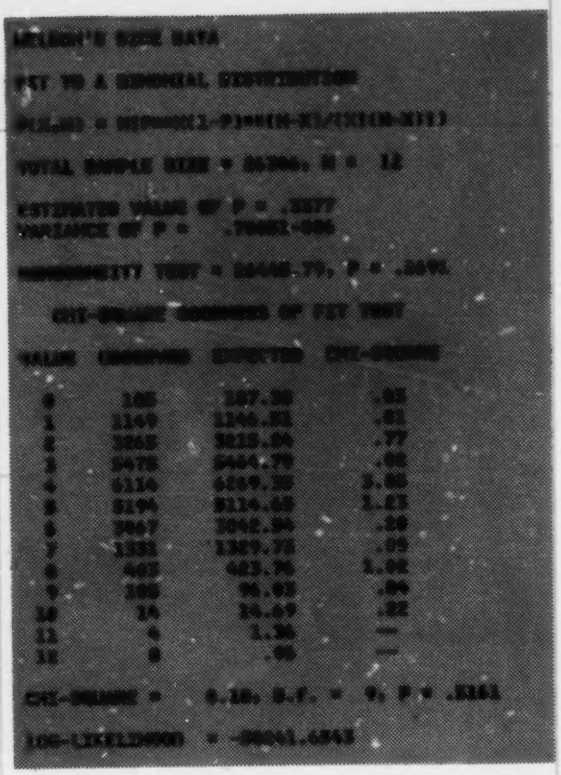


Figure 5.2

BEGIN LAST LINE OF TEXT HERE
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/3"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

6. TRUNCATED BINOMIAL DISTRIBUTION

6.1 History and Definition

The truncated binomial distribution occurs when sampling from the binomial distribution where the outcome of "0" cannot be observed:

$$\Pr[x=j] = \binom{N}{j} p^j q^{N-j}/(1-q^N) \qquad (6.1)$$

subject to $q = 1 - p$, $0 < p < 1$, and $j > 0$. The distribution was expressed in this form by Finney [11] in 1949. The distribution with just the zeros truncated has also been called the positive binomial distribution, since the truncation could include more than the zeros, or could occur on the right. As with the binomial distribution, we assume that the distribution (6.1) is sampled n times, so that the outcome i occurs f_i times. Thus the f_i's sum to n.

6.2 Calling Sequence

The truncated binomial distribution routine is called by

Col. 8	15	25
TBINOMAL	<N>	<P>

where N is an integer, and P is the value of p to be fitted. If N is not specified, it will be set to K, the value of the largest count. If p is not specified it will be estimated from the data. Note that the main program requires that the number of "0" counts be read in, even though it is ignored by this routine.

6.3 Maximum Likelihood Estimates

The estimate of p is calculated using the MIP (or EM algorithm) of Dempster, Laird, and Rubin [9]. The initial estimate comes from Mantel [22] in 1950.

$$\hat{p} = (\bar{x}-f_1/n)/(N-f_1/n) \qquad (6.2)$$

where $\bar{x} = \sum_{i=1}^{N} if_i/(nN)$.

The MIP iteration scheme is then computed by estimating the "missing" cell, f_0

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin.) (4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA

$$\hat{f}_0 = (\hat{q})^N n/(1-\hat{q}^N) \qquad (6.3)$$

DROPPED HEAD: BEGIN SECTIONS HERE

where $\hat{q} = 1 - \hat{p}$. The parameter, p, is then reestimated

$$\hat{\hat{p}} = \sum_{i=1}^{N} if_i(1-\hat{q}^N)/(nN). \qquad (6.4)$$

Equations 6.3 and 6.4 define the iteration scheme. The iterations are continued until the relative change in $\hat{p}$ is less than 10^{-5}.

The asymptotic variance of $\hat{p}$ is estimated by

$$Var(\hat{p}) = \hat{\hat{p}}\hat{\hat{q}}(1-\hat{q}^N)^2/[Nn(1-\hat{q}^N-N\hat{\hat{p}}\hat{\hat{q}}^{N-1})] \qquad (6.5)$$

which comes from the second partial derivatives of the likelihood function.

6.4 Test for Goodness-of-Fit

A standard goodness of fit test is computed for the observed frequencies:

$$\chi^2 = \sum_{i=1}^{K} (f_i - E_i)^2/E_i, \text{ where } E_i = nPr[x=i] \qquad (6.6)$$

and the cells are collapsed from the left and right until the end cells each have an expected value of 5. The degrees of freedom will be the number of cells after collapsing minus two. If P is specified, then only one will be subtracted.

6.5 An Example

Finney and Varley [12] investigated the distributions of gall-fly eggs and gall-cells in the production of flowerheads of black knapweed in 1935 and 1936. This example is also discussed in section 10.5. Since the maximum number of gall-fly eggs per flowerhead is about 15, and since no flowerheads are observable with no eggs, a truncated binomial with N = 15 was fitted to the data. The fit is not as good as the fit of the truncated Poisson distribution (see section 10.5).

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE

CENTER OF PAGE

TOP OF IMAGE AREA

Input for Varley's data:

DROPPED HEAD; BEGIN SECTIONS HERE

```
VARLEY'S HALL-FLY EGG COUNTS FROM 1953
  12
 (1X,I2)
   0  29   36  36  33   8   5   0   2   1   0   0   1
TRIGONAL    15
```

Output:

```
VARLEY'S HALL-FLY EGG COUNTS FROM 1953

FIT TO A TRUNCATED BINOMIAL DISTRIBUTION

TOTAL SAMPLE SIZE =   146, N =  15

ESTIMATED VALUE OF P = .1953
ASYMP. VAR. OF P =   .7923E-003

   CHI-SQUARE GOODNESS OF FIT TEST

VALUE  OBSERVED  EXPECTED  CHI-SQUARE

  1       29       23.05       1.37
  2       36       27.68        .82
  3       33       23.63        .34
  4       23       27.46        .79
  5        8       15.99       3.97
  6        5        8.28       1.21
  7        0        1.69        --
  8        2         .42        --
  9        1         .08        --
 10        0         .01        --
 11        0         .00        --
 12        1         .00        --
 13+       0         .00        --

CHI-SQUARE =   10.30, D.F. =  6, P = .0784

LOG-LIKELIHOOD =   -255.0476
```

Figure 6.1

BEGIN LAST LINE OF TEXT

BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

3/8"

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

7. MIXTURE OF TWO BINOMIAL DISTRIBUTIONS

7.1 History and Definition

If samples are taken randomly from two different binomial distributions with a common known N, then the result is a mixture of two binomial distributions. If the probability of sampling the first binomial distribution (with parameter p_1) is α, the resulting distribution is

$$\Pr[x=j] = \alpha \binom{N}{j} p_1^j q_1^{(N-j)} + (1-\alpha)\binom{N}{j} p_2^j q_2^{(N-j)}$$

$$= \alpha g_1(j) + (1-\alpha) g_2(j)$$

subject to $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$. (7.1)

This distribution was described by Rider [32] and moment estimators were given. The estimation of parameters from mixtures of distributions is difficult by any method, and usually requires large sample sizes.

7.2 Calling Sequence

The binomial mixture routine is called by

Col. 8	15	25	35	45
BINOMIX	<N>	<P_1>	<P_2>	<A>

where N is an integer, and P_1, P_2, and A are the values of p_1, p_2, and α to be fitted. If N is not specified it will be set to the value of the largest frequency, K. If any of the other parameters are not specified, they will be estimated from the data.

7.3 Maximum Likelihood Estimates

Arbitrary initial estimates are first calculated:

$$\hat{\alpha} = 1/2,\ \hat{p}_1 = (3/4)\hat{p},\ \hat{p}_2 = (4/3)\hat{p}_1 \quad (7.2)$$

where $\hat{p}$ is given by equation (5.2). The initial estimates are constrained to the interval [.05, .95]. The (MIP or EM) iteration scheme has been given by Hasselblad [17].

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.) (4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

$$\hat{\hat{p}}_1 = \sum_{i=0}^{K} \hat{\hat{\alpha}}\hat{g}_1(i)f_i i/[(\hat{\hat{\alpha}}\hat{g}_1(i)+(1-\hat{\alpha})\hat{g}_2(i))Nn] \qquad (7.3)$$

$$\hat{\hat{p}}_2 = \sum_{i=0}^{K} (1-\hat{\alpha})\hat{g}_2(i)f_i i/[(\hat{\hat{\alpha}}\hat{g}_1(i)+(1-\hat{\alpha})\hat{g}_2(i))Nn] \qquad (7.4)$$

and

$$\hat{\hat{\alpha}} = \sum_{i=0}^{K} \hat{\hat{\alpha}}\hat{g}_1(i)f_i/[(\hat{\hat{\alpha}}\hat{g}_1(i)+(1-\hat{\alpha})\hat{g}_2(i))Nn] \qquad (7.5)$$

where $\hat{g}_1(i)$ and $\hat{g}_2(i)$ are in equation (7.1), with the parameters replaced by their estimates. The estimated asymptotic covariance matrix is calculated from the second partial derivatives of the likelihood function, which are given by Hasselblad [17].

7.4 Test for Goodness-of-Fit

A standard goodness-of-fit test is computed for the observed frequencies:

$$X^2 = \sum_{i=0}^{K} (f_i - E_i)^2/E_i, \text{ where } E_i = nPr[x=i] \qquad (7.6)$$

and the cells are collapsed from the left and right until the end cells each have an expected value of 5. The degrees of freedom will be the number of cells after collapsing minus four. If any of the parameters are specified, the degrees of freedom will be adjusted.

7.5 An Example

Sokal and Rohlf [36] give the observed frequencies of the number of males in 6115 sibships of size 12. The data are from Saxony from 1876 to 1885. A mixture of two binomial distributions gives a reasonable fit ($p=.0898$) to the data, whereas a single binomial gives a very poor fit ($X^2 = 105.79$, D.F. = 9, $p<.0001$). A likelihood ratio test of the difference between the two distributions gives an asymptotic X^2 of -2(-12534.17 + 12492.54) = 83.26 for 2 degrees of freedom ($p<.001$).

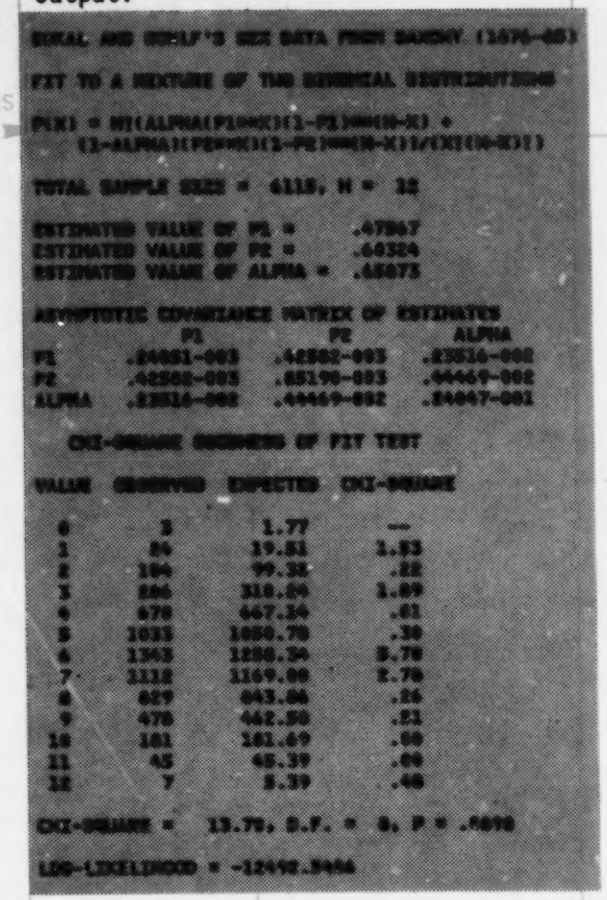
Input for Sokal and Rohlf's data:

```
SOKAL AND ROHLF'S SEX DATA FROM SAXONY (1876-85)
   12
(13I5)
    3   24  104  286  670 1033 1343 1112  829  478  181   45    7
BINOMIX      12
```

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA. OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

Output:



```
RURAL AND URBAN'S SIZE DATA FROM SAMONY (1970-85)

FIT TO A MIXTURE OF TWO BINOMIAL DISTRIBUTIONS

P(X) = N!(ALPHA(P1**X)(1-P1)**(N-X) +
    (1-ALPHA)(P2**X)(1-P2)**(N-X))/(X!(N-X)!)

TOTAL SAMPLE SIZE =  6118, N =  12

ESTIMATED VALUE OF P1 =      .47867
ESTIMATED VALUE OF P2 =      .66324
ESTIMATED VALUE OF ALPHA =   .45873

ASYMPTOTIC COVARIANCE MATRIX OF ESTIMATES
              P1              P2            ALPHA
P1       .84681-003      .42562-003     .83516-002
P2       .42562-003      .85196-003     .94469-002
ALPHA    .83516-002      .94469-002     .84047-001

   CHI-SQUARE GOODNESS OF FIT TEST

VALUE  OBSERVED  EXPECTED  CHI-SQUARE

  0        3       1.77        --
  1       84      19.81       1.83
  2      184      99.31        .22
  3      886     318.84       1.89
  4      670     667.34        .01
  5     1033    1050.78        .30
  6     1343    1288.34       8.78
  7     1118    1169.08       2.78
  8      629     643.06        .16
  9      470     442.50        .21
 10      101     101.69        .00
 11       45      15.39        .08
 12        7       0.39        .48

CHI-SQUARE =   13.78, D.F. =  8, P = .0898

LOG-LIKELIHOOD = -12492.3466
```

Figure 7.1

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

8. THE BETA-BINOMIAL DISTRIBUTION

8.1 History and Definition

If the sampling occurs from a binomial distribution where the parameter itself is the outcome of a beta distribution, then the resulting distribution is a beta-binomial distribution. This distribution was described by Skellam [35] and can be written

$$Pr[x=j] = \binom{N}{j} \frac{B(\alpha+j,\beta+N-j)}{B(\alpha,\beta)}, \quad j=0,1,2,\ldots,N \qquad (8.1)$$

where $B(\alpha,\beta)$ is the beta function.

Equivalently

$$Pr[x=j] = \frac{1}{C} \frac{\alpha(\alpha+1)\ldots(\alpha+j-1)}{j!} \quad \frac{\beta(\beta+1)\ldots(\beta+N-j-1)}{(N-j)!},$$

where $C = \binom{\alpha+\beta+N-1}{N}$. (8.2)

8.2 Calling Sequence

The beta-binomial distribution routine is called by

Col. 8	15	25	35
BETBINOM	<N>	<A>	<B>

N is the parameter in (8.1), read in as an integer, and if not specified is set to K, the value of the largest count. A and B are the values of α and β to be fitted. If either or both are not given, they are estimated from the data.

8.3 Maximum Likelihood Estimates

The estimates of α and β are calculated using a modified Gauss-Newton method. The initial estimates are computed from the first two factorial moments, R_1 and R_2:

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

$$R_1 = \sum_{i=0}^{N} if_i/n \qquad (8.3)$$

$$R_2 = \sum_{i=1}^{N} i(i-1)f_i/(nR_1). \qquad (8.4)$$

The initial estimates are

$$\hat{\alpha} = (R_1R_2-(N-1)R_1)/((N-1)R_1-NR_2) \qquad (8.5)$$

$$\hat{\beta} = \hat{\alpha}(N/R_1-1). \qquad (8.6)$$

The Gauss-Newton iteration scheme requires the elements of the first partial derivatives:

$$d_{1i} = \psi(\alpha+\beta)-\psi(\alpha)+\psi(\alpha+i) \qquad (8.7)$$

$$d_{2i} = \psi(\alpha+\beta)-\psi(p)+\psi(\beta+N-i) \qquad (8.8)$$

where ψ is the derivative of the log-gamma function (digamma function) and α and β are evaluated using their previous estimates. The first partial derivatives are

$$D_j = \sum_{i=0}^{N} d_{ji}, \qquad j=1,2 \quad . \qquad (8.9)$$

The matrix of cross products is

$$C_{jk} = \sum_{i=0}^{N} d_{ji}d_{ki} \qquad j,k=1,2 \quad . \qquad (8.10)$$

The iteration scheme is

$$\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} = \begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} + C^{-1}D \quad . \qquad (8.11)$$

The estimates are constrained to be positive. If equation (8.11) does not improve the likelihood, the correction is multiplied by 1/2 until the likelihood is improved. The iteration scheme is continued until the change in α and β is less than .00001.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA. OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

8.4 Test for Goodness-of-Fit

A standard goodness-of-fit test is computed for the observed frequencies:

$$X^2 = \sum_{i=0}^{K} (f_i - E_i)^2/E_i, \text{ where } E_i = nPr[x=i] \qquad (8.12)$$

and the cells are collapsed from the left and right until the end cells each have an expected value of 5. The degrees of freedom will be the number of cells after collapsing minus three. If α or β are specified then one degree of freedom will be added for each parameter specified.

8.5 Examples

Haseman and Soares [16] fit a beta-binomial distribution to fetal mortality of mice. The N of the distribution is the number of implants. Some 524 CF1S mice were tested, where N ranged from 1 to 20. Figures 8.1 and 8.2 give the fits for N's of 12 and 13, both of which give adequate fits as measured by a chi-square test.

Input for Haseman and Soares data:

```
HASEMAN AND SOARES DATA
   12
(13I5)
   33   24   11    5    4    4    0    0    0    0    0    0    1
BETBINOM

HASEMAN AND SOARES DATA
    0
(10I5)
   39   27   12    6    5    2    0    0    1
BETBINOM     13
```

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD. BEGIN SECTIONS HERE

Outputs:

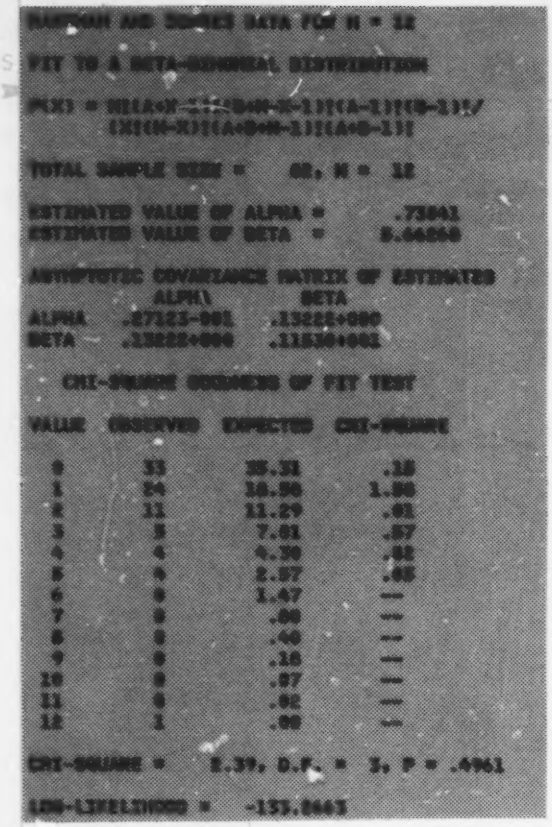


Figure 8.1

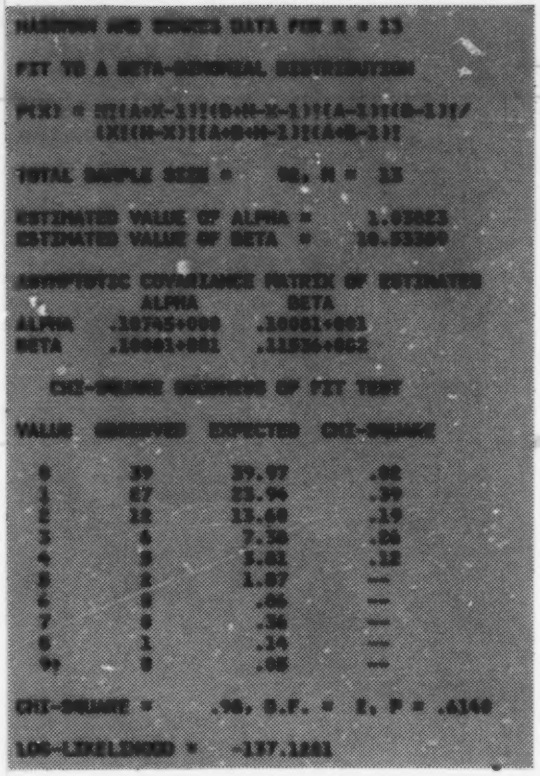


Figure 8.2

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA. OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD. BEGIN SECTIONS HERE

9. THE POISSON DISTRIBUTION

9.1 History and Definition

In 1837 Poisson derived the distribution which has since been given his name. The Poisson distribution takes on the values 0, 1, 2,...,where

$$Pr[x=j] = e^{-\lambda}\lambda^j/j! \quad \text{for } \lambda>0. \tag{9.1}$$

It is assumed that the distribution is sampled n times so that the outcome i occurs f_i times. Thus the f_i's sum to n.

9.2 Calling Sequence

The Poisson distribution routine is called by

Col. 8		25
POISSON		<L>

where L (optional) is a specified value of λ to be fitted and tested for goodness-of-fit. If not specified, λ is estimated from the data and then fitted.

9.3 Maximum Likelihood Estimate

The maximum likelihood estimate for λ is

$$\hat{\lambda} = \sum_{i=0}^{K} if_i/n \tag{9.2}$$

This estimate is also the minimum variance unbiased estimate. The variance of this estimate is estimated by $\hat{\lambda}/n$.

9.4 Tests for Goodness-of-Fit

A standard goodness-of-fit test is computed for the observed frequencies:

$$X^2 = \sum_{i=0}^{K} (f_i-E_i)^2/E_i, \tag{9.3}$$

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA. OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.) (4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

where $E_i = n \Pr[x=i]$, and the value of λ is taken to be $\hat{\lambda}$ unless specified by L. The cells are collapsed from the left and right until the end cells each have an expected value of 5. The degrees of freedom will be the number of cells after collapsing minus two.

A second goodness-of-fit test is commonly known as the homogeneity test. It is a test of the variance of the sample, and is discussed by Potthoff and Whittinghill [31]. If λ is estimated, then

$$\chi^2 = \sum_{i=0}^{K} f_i i^2/\hat{\lambda} - n\hat{\lambda} \qquad (9.4)$$

is approximately χ^2 with n-1 degrees of freedom. If λ is specified, the quantity

$$\chi^2 = \sum_{i=0}^{K} f_i i^2/\lambda - 2 \sum_{i=0}^{K} f_i i/n + n\lambda \qquad (9.5)$$

is approximately χ^2 with n degrees of freedom.

9.5 An Example

Mutagenicity testing using the Ames test [2] has recently received a great deal of interest. The test consists of counts of revertants per plate for each dose of a given compound. The following example is the result of 196 tests at a zero dose (controls) for the strain TA1537 without metabolic activation. The plates were all made on the same day. After a 48-hour incubation period revertants per plate were counted manually. The fit is adequate as measured by both the variance test and the standard chi-square goodness-of-fit test.

Input for TA1537 Ames test data:

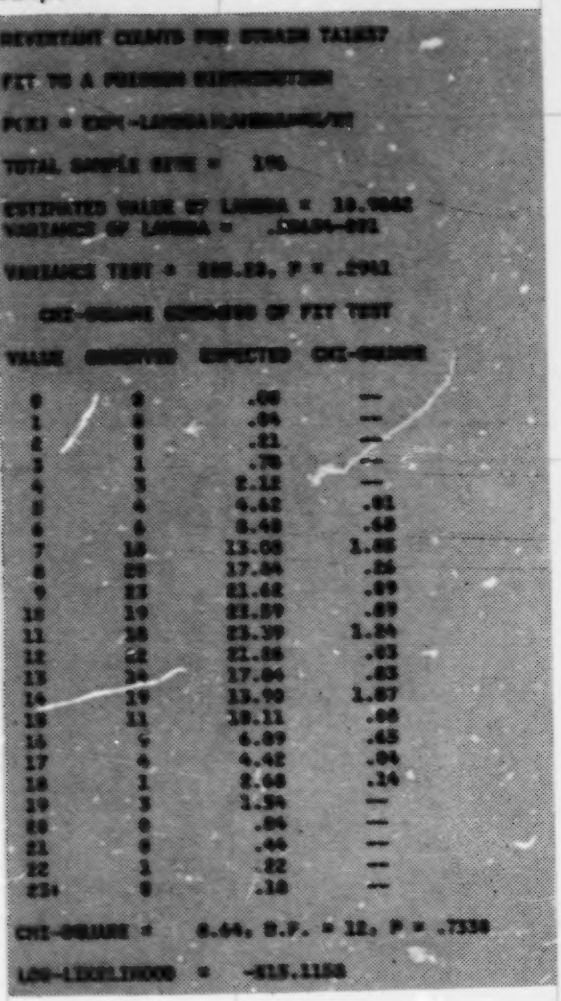
```
REVERTANT COUNTS FOR STRAIN TA1537
    22
(23I3)
  0  0  0  1  3  4  6 18 20 23 19 18 22 14 19 11  9  4  1  3  0  0  1
POISSON
```

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

Output:



REVERTANT COLONIES PER STRAIN TA1537

FIT TO A POISSON DISTRIBUTION

PCR1 = EXP(-LAMBDA)LAMBDA**R/R!

TOTAL SAMPLE SIZE = 15

ESTIMATED VALUE OF LAMBDA = 10.7667
VARIANCE OF LAMBDA = .53834-001

VARIANCE TEST = 388.89, P = .2961

CHI-SQUARE GOODNESS OF FIT TEST

VALUE	OBSERVED	EXPECTED	CHI-SQUARE
0	0	.00	—
1	0	.04	—
2	0	.21	—
3	1	.78	—
4	3	2.12	—
5	4	4.62	.81
6	6	8.48	.68
7	10	13.05	1.88
8	20	17.84	.26
9	23	21.62	.09
10	19	23.29	.89
11	28	23.39	1.24
12	22	21.86	.03
13	24	17.84	.83
14	19	13.99	1.87
15	11	10.11	.08
16	7	6.89	.45
17	4	4.42	.04
18	1	2.68	.14
19	3	1.54	—
20	0	.84	—
21	0	.44	—
22	1	.22	—
23+	0	.18	—

CHI-SQUARE = 8.44, D.F. = 12, P = .7338

LOG-LIKELIHOOD = -615.1158

Figure 9.1

BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
BEGIN LAST LINE OF TEXT
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

10. TRUNCATED POISSON DISTRIBUTION

10.1 History and Definition

The truncated Poisson distribution occurs when sampling from a Poisson distribution where the "0" outcome cannot be observed:

$$Pr[x=j] = e^{-\lambda}\lambda^{j}/(j!(1-e^{-\lambda})),\ j = 1,2,3,... \qquad (10.1)$$

The distribution was described as early as 1926 by McKendrick [23]. The distribution has also been referred to as the positive Poisson distribution. It is assumed that the distribution is sampled n times and that the outcome i occurs f_i times. Thus the f_i's ($i = 1, 2, 3,...$) sum to n.

10.2 Calling Sequence

The truncated Poisson distribution is called by

Col. 8	25
TPOISSON	<L>

where L (optional) is a specified value of λ to be fitted and tested for goodness-of-fit. If not specified, λ is estimated from the data, and then fitted. Note that the main program requires that the number of "0" counts be read in, even though it is ignored by this routine.

10.3 Maximum Likelihood Estimates

The estimate of λ is calculated using the MIP (or EM alogirthm). The missing information in this case is f_0 (the number of 0's). The initial estimate was provided by Irwin [18] who ascribed it to McKendrick [23].

$$\hat{f}_0 = \left(\sum_{i=1}^{K} i\, f_i\right)^2 / \left(\sum_{i=1}^{K} i(i-1) f_i\right) - n \qquad (10.2)$$

The initial estimate of λ then becomes

$$\hat{\lambda} = \left(\sum_{i=1}^{K} i\, f_i\right) / (n+\hat{f}_0) \qquad (10.3)$$

BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
BEGIN LAST LINE OF TEXT
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

The iteration scheme is to reestimate f_0

$$\hat{\hat{f}}_0 = e^{-\hat{\lambda}} n/(1-e^{-\hat{\lambda}}), \qquad (10.4)$$

and then λ can be reestimated using equation (10.2). The asymptotic variance of $\hat{\lambda}$, as given by Finney and Varley [12], is

$$\mathrm{Var}(\hat{\lambda}) = \lambda^2/(n\bar{x}(\lambda+1-\bar{x})) \qquad (10.5)$$

where $\bar{x}$ is the arithmetic mean of the observed counts.

10.4 Test for Goodness-of-Fit

A standard goodness of fit test is computed for the observed frequencies:

$$X^2 = \sum_{i=1}^{K} (f_i - E_i)^2/E_i, \text{ where } E_i = nPr[x=i] \qquad (10.6)$$

and the cells are collapsed from the left and right until the end cells each have an expected value of 5. The degrees of freedom will be the number of cells after collapsing minus two. If λ is specified, then only one will be subtracted.

10.5 An Example

Finney and Varley [12] investigated the distributions of gall-fly eggs and gall-cells in the production of flowerheads of black knapweed in 1935 and 1936. This is the same data set considered by Irwin [18]. Figures 10.1 and 10.2 show the fitted and observed frequencies of flowerheads in 1935 and 1936 respectively. A fit of the 1935 and 1936 data combined gives a log-likelihood of -443.1378. The likelihood ratio test for identical λ's in each year is -2(-443.1378 + 276.5876 + 166.5481) = .004. This value is approximately distributed as a chi-square with one degree of freedom, indicating that there was no difference in the two years.

Input for Varley data:

```
VARLEY'S GALL-FLY EGG COUNTS FROM 1935
   12
 (13I5)
    0   29   38   36   23    8    5    3    2    1    0    0    1
TPOISSON

VARLEY'S GALL-FLY EGG COUNTS FROM 1936
   12
(13I5)
    0   22   18   18   11    9    6    3    0    1    0    0    0
TPOISSON
```

BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
BEGIN LAST LINE OF TEXT
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD, BEGIN SECTION HERE

Outputs:

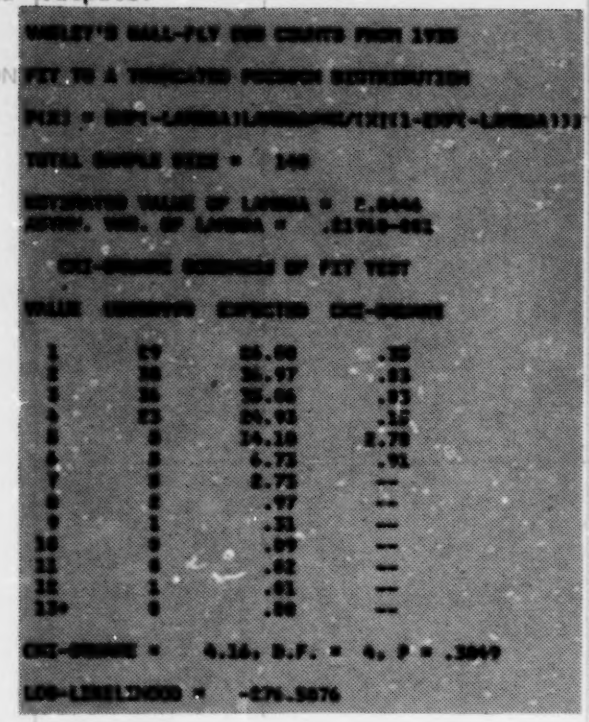


Figure 10.1

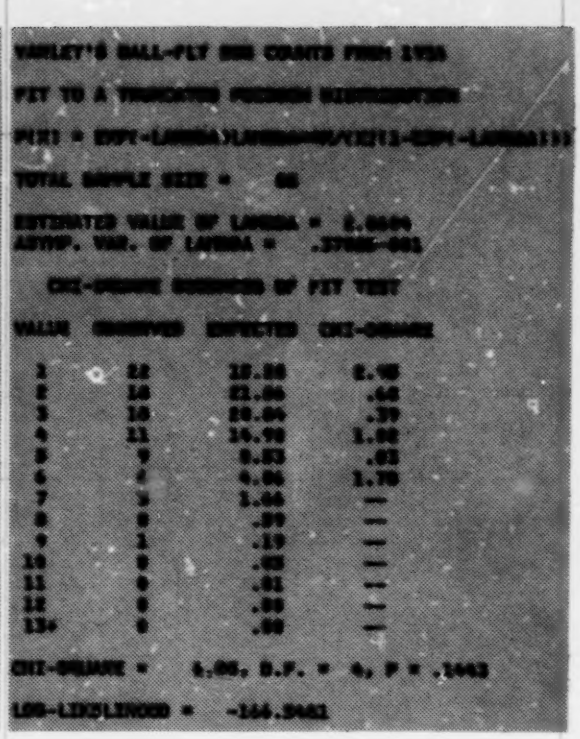


Figure 10.2

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

11. MIXTURE OF TWO POISSON DISTRIBUTIONS

11.1 History and Definition

If samples are taken randomly from two different Poisson distributions the result is a mixture of two Poisson distributions. If the probability of sampling the first Poisson distribution (with parameter λ_1) is α, the resulting distribution is

$$\Pr[x=j] = \alpha \; e^{-\lambda_1}\lambda_1^j/j! + (1-\alpha)\, e^{-\lambda_2}\lambda_2^j/j! \qquad (11.1)$$

$$= \alpha g_1(j) + (1-\alpha) g_2(j)$$

This distribution was described by Schilling [34] and moment estimators were given. The estimation of parameters from mixtures of distributions is difficult by any method, and usually requires large sample sizes.

11.2 Calling Sequence

The Poisson mixture routine is called by

Col. 8	25	35	45
POISSMIX	<L_1>	<L_2>	<A>

where L_1, L_2 and A are the values of λ_1, λ_2, and α to be fitted. If any of the parameters are not specified, they will be estimated from the data.

11.3 Maximum Likelihood Estimates

Arbitrary initial estimates are first calculated:

$$\hat{\lambda}_1 = (3/4)\hat{\lambda} \quad \text{and} \quad \hat{\lambda}_2 = (4/3)\hat{\lambda} \qquad (11.2)$$

where $\hat{\lambda}$ is given by equation (9.2). The (MIP or EM) iteration scheme has been given by Hasselblad [17].

BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
BEGIN LAST LINE OF TEXT
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

$$\hat{\lambda}_1 = \sum_{i=0}^{K} \hat{\alpha}\hat{g}_1(i)f_i i/[(\hat{\alpha}\hat{g}_1(i)+(1-\hat{\alpha})\hat{g}_2(i))n] \qquad (11.3)$$

$$\hat{\lambda}_2 = \sum_{i=0}^{K} (1-\hat{\alpha})\hat{g}_2(i)f_i i/[(\hat{\alpha}\hat{g}_1(i)+(1-\hat{\alpha})\hat{g}_2(i))n] \qquad (11.4)$$

and

$$\hat{\alpha} = \sum_{i=0}^{K} \hat{\alpha}\hat{g}_1(i)f_i/[(\hat{\alpha}\hat{g}_1(i)+(1-\hat{\alpha})\hat{g}_2(i))n] \qquad (11.5)$$

where $\hat{g}_1(i)$ and $\hat{g}_2(i)$ are in equation (11.1), with the parameters replaced by their estimates. The estimated asymptotic covariance matrix is calculated from the second partial derivatives of the likelihood function.

11.4 Test for Goodness-of-Fit

A standard goodness-of-fit test is computed for the observed frequencies:

$$X^2 = \sum_{i=0}^{K} (f_i - E_i)^2/E_i, \text{ where } E_i = nPr[x=i] \qquad (11.6)$$

and the cells are collapsed from the left and right until the end cells each have an expected value of 5. The degrees of freedom will be the number of cells after collapsing minus four. If any of the parameters are specified, the degrees of freedom will be adjusted.

11.5 An Example

Additional Ames test data similar to that in section 9.5 was collected by pooling the controls from several experiments. This data was fitted to a mixture of two Poisson distributions, since the variance was too large to fit a single Poisson distribution. Although it is hard to justify this biologically, the mixture does provide an adequate fit as measured by the standard chi-square goodness-of-fit test. The results are in Figure 11.1.

Input for TA1537 Ames test data:

```
COUNTS OF REVERTANTS: TA1537-
   28
(2613)
  0  1  6 13 21 24 22 23 22 28 20  6 17 11 11  4  5  5  4  0  0  1  0  2  0  0
  0  1  1
POISSMIX
```

BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
BEGIN LAST LINE OF TEXT
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

Output:

```
COUNTS OF NEGATIVES= 241537-

FIT TO A MIXTURE OF TWO POISSON DISTRIBUTIONS

P(X) = ALPHA*EXP(-LAMBDA1)*LAMBDA1**X/X! +
       (1-ALPHA)*EXP(-LAMBDA2)*LAMBDA2**X/X!

TOTAL SAMPLE SIZE =   230

ESTIMATED VALUE OF LAMBDA1 =     6.49579
ESTIMATED VALUE OF LAMBDA2 =    13.28831
ESTIMATED VALUE OF ALPHA =       .66377

ASYMPTOTIC COVARIANCE MATRIX OF ESTIMATES
            LAMBDA1         LAMBDA2          ALPHA
LAMBDA1    .13E+0**0       .36212+000      .17625-001
LAMBDA2    .18212+000      .64116+000      .39904-001
ALPHA      .17625-001      .39904-001      .45459-002

   CHI-SQUARE GOODNESS OF FIT TEST

VALUE  OBSERVED  EXPECTED  CHI-SQUARE

  0        0        .02        --
  1        1       1.60        --
  2        4       5.19        .40
  3       13      11.27        .26
  4       21      18.29        .37
  5       24      24.12        .00
  6       22      26.41        .83
  7       23      25.64        .27
  8       22      22.40        .01
  9       23      18.49        .12
 10       20      13.81       1.60
 11        6      10.65       1.49
 12       17      10.97       3.32
 13       11       9.70        .18
 14       11       8.80        .76
 15        4       7.22       1.44
 16        8       5.87        .13
 17        3       4.86        .32
 18        4       3.96        .13
 19        0       3.23        .32
 20        0       1.94        --
 21        1        .98        --
 22        0        .89        --
 23        1        .34        --
 24        0        .19        --
 25        0        .10        --
 26        0        .05        --
 27        1        .03        --
 28        1        .01        --
 29+       0        .01        --

CHI-SQUARE =   13.60, D.F. = 14, P = .4842

LOG-LIKELIHOOD =   -670.0049
```

Figure 11.1

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

12. NEGATIVE BINOMIAL DISTRIBUTION

12.1 History and Definition

The negative binomial distribution was discussed by Pascal [99] and Fermat. The earliest known publication was by Montmort [24] in 1714. An excellent review of the distribution was given by Bartko [4]. The distribution can be written in several forms, but we will use the following one

$$Pr[x=j] = \binom{N+j-1}{N-1} p^j (1+p)^{-(N+j)} \qquad j=0,1,2,... \qquad (12.1)$$

where $p>0$ and $N>0$. For the special case where N is an integer, the distribution is sometimes known as the Pascal distribution. Unlike the binomial distribution, both N and p are parameters to be estimated.

For the case of $N = 1$, the distribution is known as the geometric distribution. This may be fitted by specifying $N = 1$ on the control card.

12.2 Calling Sequence

The negative binomial distribution routine is called by

Col. 8	25	35
NEGBINOM	<P>	<N>

where P and N are the (optional) values of the parameters p and N to be fitted. If either or both are left blank, the parameters will be estimated from the data.

12.3 Maximum Likelihood Estimates

Initial estimates can be obtained by the method of moments:

$$\hat{p} = s^2/\bar{x} - 1 \qquad (12.2)$$

$$\hat{N} = \bar{x}^2/(s^2-\bar{x}^2) \qquad (12.3)$$

BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
BEGIN LAST LINE OF TEXT
3/8"

EPA-287 (Cin.) (4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

where $\bar{x}$ is the sample mean and s^2 is the sample variance. Maximum likelihood estimates can be obtained by rearranging the first partial derivatives of the log-likelihood function,

$$\log(1+\bar{x}/\hat{N}) - \sum_{i=1}^{K} S_i = 0, \qquad (12.4)$$

where $S_i = \sum_{j=i}^{K} (\hat{N}+j-1)^{-1} f_j$.

This equation involving just N can be solved with a Newton-Raphson scheme. Denote the left hand side of (12.2) by F; then

$$\frac{\partial F}{\partial N} = -(n\bar{x})/(\hat{N}^2+\hat{N}\bar{x}) + \sum_{i=0}^{K} S_i' \qquad (12.5)$$

where $S_i' = \sum_{j=i}^{K} (\hat{N}+j-1)^{-2} f_i$.

The iteration scheme is

$$N = \hat{N} - F/\frac{\partial F}{\partial N} \quad . \qquad (12.6)$$

This procedure is continued until the change in $\hat{N}$ is less than .00001. If this does not occur within 100 iterations, the routine indicates that the procedure did not converge. The estimate for p is then

$$\hat{p} = \bar{x}/\hat{N} \quad . \qquad (12.7)$$

The equations must be modified slightly if p is specified. If N is specified, then just equation (12.7) is used.

The asymptotic variances of $\hat{p}$ and $\hat{N}$ are estimated from the inverse of the second partial derivatives of the log-likelihood function,

$$\frac{\partial^2 L}{\partial p^2} = -n(Np^2-2p\bar{x}-\bar{x})/(p(1+p)), \qquad (12.8)$$

$$\frac{\partial^2 L}{\partial p \partial N} = n/(1+p), \text{ and} \qquad (12.9)$$

$$\frac{\partial^2 L}{\partial N^2} = \sum_{i=1}^{K} S_i' \quad . \qquad (12.10)$$

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD, BEGIN SECTIONS HERE

The two integer values of N closest to $\hat{N}$ are tried for the integer (Pascal) solution to the problem. The integer $\hat{N}$ giving the largest log-likelihood is given as an alternative solution.

12.4 Test for Goodness-of-Fit

A standard goodness-of-fit test is computed for the observed frequencies:

$$X^2 = \sum_{i=1}^{K} (f_i - E_i)^2 / E_i, \text{ where } E_i = nPr[x=i] \qquad (12.11)$$

and the cells are collapsed from the left and right until the end cells each have an expected value of 5. The degrees of freedom will be the number of cells after collapsing minus three. If p or N are specified, then one degree of freedom will be added for each parameter specified.

6-1/2"

12.5 Example

Love, et al. [21] measured the incidence of acute respiratory disease in children. Counts of illnesses in children aged 6 to 12 from two communities in New York are shown in Figures 12.1 and 12.2. Both fits are adequate as measured by a standard chi-square goodness-of-fit test. A fit of the two communities combined gives a log-likelihood of -598.2169. The likelihood ratio test for similar distributions in each community is -2(-598.2169 + 288.8186 + 304.1975) = 10.4. This value is approximately distributed as a chi-square with 2 degrees of freedom, and therefore indicates that the distribution of illnesses is not similar in the two communities.

Input for Love's illness data:

```
LOVE'S UPPER RESPIRATORY ILLNESSES FROM RIVERHEAD
    7
(8I5)
  185   63   28    4    5    3    1
NEGBINOM

LOVE'S UPPER RESPIRATORY ILLNESSES FROM BRONX
    7
(8I5)
  159   91   25    4    2    1    0    0
NEGBINOM
```

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTION HERE

Outputs:

```
LOVE'S UPPER RESPIRATORY ILLNESSES FROM RIVERMEAD

FIT TO A NEGATIVE BINOMIAL DISTRIBUTION

P(X) = (N+X-1)!P**N*Q**X/(X!(N-1)!(1+P)**(N+X))

TOTAL SAMPLE SIZE =   280

ESTIMATED VALUE OF P =      .76186
ESTIMATED VALUE OF N =      .80289

ASYMPTOTIC COVARIANCE MATRIX OF ESTIMATES

            P              N
P     .42318-001    -.39863-001
N    -.39863-001     .43114-001

SOLUTION FOR INTEGER N (PASCAL DISTRIBUTION):
P =  .95516, N =    1, LOG-LIKELIHOOD =   -384.83

   CHI-SQUARE GOODNESS OF FIT TEST

VALUE  OBSERVED  EXPECTED  CHI-SQUARE

  0      185      185.14       .00
  1       63       63.28       .00
  2       25       24.28       .57
  3        4        9.66      3.32
  4        3        3.91       .05
  5        3        1.68        --
  6        1         .64        --
  7        0         .27        --
  8+       0         .19        --

CHI-SQUARE =    4.73, D.F. =  2, P = .0939

LOG-LIKELIHOOD =   -384.1975
```

Figure 12.1

```
LOVE'S UPPER RESPIRATORY ILLNESSES FROM BROCK

FIT TO A NEGATIVE BINOMIAL DISTRIBUTION

P(X) = (N+X-1)!P**N*Q**X/(X!(N-1)!(1+P)**(N+X))

TOTAL SAMPLE SIZE =   280

ESTIMATED VALUE OF P =      .19478
ESTIMATED VALUE OF N =     3.61011

ASYMPTOTIC COVARIANCE MATRIX OF ESTIMATES

            P              N
P     .89182-002    -.47427+000
N    -.47427+000     .25430+002

SOLUTION FOR INTEGER N (PASCAL DISTRIBUTION):
P =  .89011, N =    4, LOG-LIKELIHOOD =   -288.82

   CHI-SQUARE GOODNESS OF FIT TEST

VALUE  OBSERVED  EXPECTED  CHI-SQUARE

  0      159      161.11       .03
  1       91       88.84       .11
  2       25       26.96       .14
  3        4        6.49       .15
  4        2        1.33        --
  5        1         .24        --
  6        0         .04        --
  7        0         .01        --
  8+       0         .00        --

CHI-SQUARE =     .63, D.F. =  1, P = .4283

LOG-LIKELIHOOD =   -288.8186
```

Figure 12.2

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD BEGIN SECTIONS HERE

13. THE TRUNCATED NEGATIVE BINOMIAL DISTRIBUTION

13.1 History and Definition

The truncated negative binomial distribution occurs when sampling from a negative binomial distribution where the "0" outcome cannot be observed:

$$Pr[x=j] = \binom{N+j-1}{N-1} p^j (1+p)^{-(N+j)} / (1-(1+p)^{-N}) \qquad j=1,2,3,... \qquad (13.1)$$

where $p>0$ and $N>0$. Examples of this distribution were given by Sampford [33] and Brass [7]. The article by Sampford gives the maximum likelihood equations. Wyshak [37] actually gives a computer subroutine to estimate the parameters N and w, where $w = 1/(1+p)$. The subroutine also calculates the estimated asymptotic variances of $\hat{N}$ and $\hat{w}$. Our program modified these equations to obtain $\hat{N}$ and $\hat{p}$, along with their estimated variances.

13.2 Calling Sequence

The truncated negative binomial distribution is called by

Col. 8	25	35
TNEGBINM	<P>	<N>

where P and N are the (optional) values of the parameters p and N to be fitted. If either or both are left blank, the parameters will be estimated from the data. Note that the main program requires that the number of "0" counts be read in, even though it is ignored by this routine.

13.3 Maximum Likelihood Estimates

Wyshak [37] gives a FORTRAN algorithm for obtaining maximum likelihood estimates using a Newton-Raphson method. For simplicity, the parameter p is replaced by w where $w = 1/(1+p)$. Initial estimates for w and N were given by Brass [7]:

$$\hat{w} = \bar{x}/s^2(1-f_0/n), \qquad (13.2)$$

and

$$\hat{N} = (\hat{w}\bar{x}-f_0/n)/(1-\hat{w}) \qquad (13.3)$$

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin.) (4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

where $\bar{x}$ is the sample mean and s^2 is the sample variance. The first and second partial derivatives of the log-likelihood were given by Wyshak and are not repeated here. The covariance terms for p are obtained from those of w using standard calculus formulas.

The iteration scheme is continued until the first derivatives with respect to $\hat{w}$ and $\hat{N}$ are less than .0001. If this does not occur within 100 iterations, the routine indicates that the procedure did not converge. This routine can fail to converge if the variance of the data is too small. In such cases, the truncated Poisson distribution may provide a good fit to the data.

13.4 Test for Goodness-of-Fit

A standard goodness-of-fit test is computed for the observed frequencies:

$$X^2 = \sum_{i=1}^{K} (f_i - E_i)^2/E_i, \text{ where } E_i = nPr[x=i] \qquad (13.4)$$

and the cells are collapsed from the left and right until the end cells each have an expected value of 5. The degrees of freedom will be the number of cells after collapsing minus three. If p or N are specified, then one degree of freedom will be added for each parameter specified.

13.5 Example

Sampford [33] gives counts of chromosome breaks. Figure 13.1 shows the fit of a truncated negative binomial distribution to this data. The standard chi-square goodness-of-fit test shows that the data are adequately fitted by this distribution. The data can also be described by a logarithm distribution (see section 14.5).

Input for Sampford's chromosome break data:

```
SAMPFORD'S CHROMOSOME BREAKS PER CELL
   13
(14I5)
    0   11    6    4    5    0    1    0    2    1    0    1    0    1
THEOBINM
```

BEGIN LAST LINE OF TEXT HERE
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-297 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

Output:

```
LUDYUNG'S CONTRIBUTIONS BREAKS PER CELL

FIT TO A TRUNCATED NEGATIVE BINOMIAL

P(X) = (N+X-1)!P**X/(X!(N-1)!(1+P)**(N+X)(1-1/(1+P)**N))

TOTAL SAMPLE SIZE =    32

ESTIMATED VALUE OF P =      3.73048
ESTIMATED VALUE OF N =       .49346

ASYMPTOTIC COVARIANCE MATRIX OF ESTIMATES

              P              N
P      .69918+001    -.18868+002
N     -.18868+001     .27667+000

    CHI-SQUARE GOODNESS OF FIT TEST

VALUE  OBSERVED  EXPECTED  CHI-SQUARE

  1        11      10.60        .08
  2         6       6.36        .02
  3         4       4.17        .01
  4         5       2.87       1.36
  5         0       2.03       2.03
  6         1       1.47        .01
  7         0       1.07         --
  8         2        .79         --
  9         1        .59         --
 10         0        .44         --
 11         1        .33         --
 12         0        .25         --
 13         1        .19         --
 14+        0        .62         --

CHI-SQUARE =    3.65, D.F. =  3, P = .3014

LOG-LIKELIHOOD =    -66.0882
```

Figure 13.1

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cln.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD. BEGIN SECTIONS HERE

14. THE LOGARITHMIC DISTRIBUTION

14.1 History and Definition

The logarithmic distribution (more formally referred to as the logarithmic series distribution) is a single parameter distribution. The distribution takes on the values 1, 2, 3,..., where

$$Pr[x=j] = -\theta^x/(x\log(1-\theta)) \text{ for } 0<\theta<1 \ . \qquad (14.1)$$

The name of the distribution comes from the series expansion of $\log(1-\theta)$. Nelson and David [26] published an excellent review paper on the distribution. The distribution itself was first described by Fisher, Corbet, and Williams [14]. The distribution can be thought of as a limiting distribution of the truncated negative binomial where the parameter N approaches zero.

It is assumed that the distribution (14.1) is sampled n times so that the outcome i occurs f_i times. Thus the f_i's (i=1, 2, ...,K) sum to n.

14.2 Calling Sequence

The logarithmic distribution routine is called by

Col. 8	25
LOGARITH	<T>

where T (optional) is a specified value of θ to be fitted and tested for goodness-of-fit. If not specified, θ is estimated from the data and then fitted. Note that the main program requires that the number of "0" counts be read in, even though it is ignored by this routine.

14.3 Maximum Likelihood Estimates

The estimate of θ is calculated using a Newton-Raphson iteration scheme. The initial estimate comes from Birch [5]:

$$\hat{\theta} = 1 - (1+[(5/3-\log\bar{x}/16)(\bar{x}-1)+2]\log\bar{x})^{-1} \ . \qquad (14.2)$$

The Newton-Raphson iteration scheme is

$$\hat{\hat{\theta}} = \hat{\theta} - (\bar{x}+\hat{\theta}/((1-\hat{\theta})\log(1-\hat{\theta})))(((1-\hat{\theta})^2\log^2(1-\hat{\theta})/(\log(1-\hat{\theta})+\theta))) \ . \qquad (14.3)$$

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA. OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cln.) (4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD. BEGIN SECTIONS HERE

The iteration scheme is continued until the change in $\hat{\theta}$ is less than .00001. The asymptotic variance of $\hat{\theta}$ is estimated by

$$\mathrm{Var}(\hat{\theta}) = \hat{\theta}^2/[n\bar{x}^2(1/(1-\theta)-\bar{x})] \quad . \tag{14.4}$$

14.4 Test for Goodness-of-fit

A standard goodness-of-fit test is computed for the observed frequencies:

$$X^2 = \sum_{i=1}^{K} (f_i - E_i)^2/E_i, \tag{14.5}$$

where $E_i = n\ \Pr[x=i]$, and the value of θ is taken to be $\hat{\theta}$ unless specified by T. The cells are collapsed from the left and right until the end cells each have an expected value of 5. The degrees of freedom will be the number of cells after collapsing minus two. If θ is specified, an additional degree of freedom will be added.

14.5 Example

The data of Sampford [33] given in section 13.5 can also be fitted by a logarithmic distribution. The output is shown in Figure 14.1.

Input for Sampford's chromosome break data:

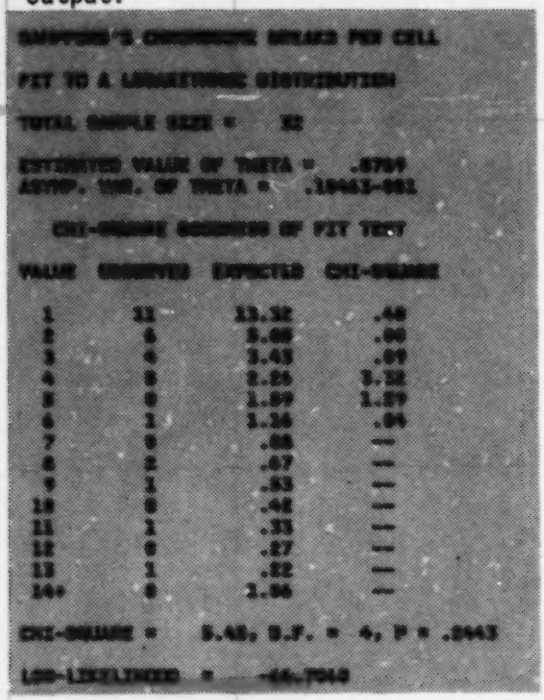
```
SAMPFORD'S CHROMOSOME BREAKS PER CELL
   13
(14I5)
    0   11    6    4    5    0    1    0    2    1    0    1    0    1
LOGARITH
```

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD. BEGIN SECTIONS HERE

Output:



```
SAMPSON'S CHROMOSOME BREAKS PER CELL

FIT TO A LOGARITHMIC DISTRIBUTION

TOTAL SAMPLE SIZE =    32

ESTIMATED VALUE OF THETA =    .8789
ASYMP. VAR. OF THETA =    .19463-001

   CHI-SQUARE GOODNESS OF FIT TEST

VALUE  OBSERVED  EXPECTED  CHI-SQUARE

  1       11      13.32       .48
  2        6       5.88       .00
  3        4       3.43       .09
  4        0       1.51      1.51
  5        0       1.07      1.07
  6        1       1.16       .04
  7        0        .88         -
  8        2        .67         -
  9        1        .53         -
 10        0        .42         -
 11        1        .33         -
 12        0        .27         -
 13        1        .22         -
 14+       0       1.36         -

CHI-SQUARE =    3.48, D.F. =  4, P = .2443

LOG-LIKELIHOOD =    -44.7940
```

Figure 14.1

BEGIN LAST LINE OF TEXT
3/8"
BOTTOM OF IMAGE AREA. OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

15. REFERENCES

1. Alldredge, J. R., and D. D. Bolstad (1977). GDIST: A Computer Code for Analysis of Statistical Distributions of Physical Data, United States Department of Commerce, NTIS TN33.071, 57 pp.

2. Ames, B. N., J. McCann, and E. Yamasaki (1975). Methods for Detecting Carcinogens and Mutagens with the Salmonella/Mammalian-Microsome Mutagenicity Test, Mutation Research, 31, pp 347-364.

3. Barr, A. J., J. H. Goodnight, J. P. Sall, and J. T. Helwig (1976). A User's Guide to SAS 76, SAS Institute, Incorporated, Raleigh, North Carolina, 330 pp.

4. Bartko, J. J. (1961). The Negative Binomial Distribution: A Review of Properties and Applications, The Virginia Journal of Science (New Series), 12, pp 18-37.

5. Birch, M. W. (1963). An Algorithm for the Logarithmic Series Distribution, Biometrics, 19, pp 651-653.

6. Bouver, H., and R. E. Bargmann (1975). Computational Algorithms and Modules for the Evaluation of Statistical Distribution Functions, Univ. of Georgia, Dept. of Statistics, Athens, Georgia.

7. Brass, W. (1958). Simplified Methods of Fitting the Truncated Negative Binomial Distribution, Biometrika, 45, pp 59-68.

8. Day, N. E. (1969). Estimating the Components of a Mixture of Normal Distributions, Biometrika, 56, 3, pp 463-484.

9. Dempster, A. D., N. M. Laird, and D. B. Rubin (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm, Journal of the Royal Statistical Society B, 38, pp 1-22.

10. Dixon, W. J. (1975). Biomedical Medical Programs (BMD-P), University of California Press, Berkeley, California, 792 pp.

11. Finney, D. J. (1949). The Truncated Binomial Distribution, Annals of Eugenics, London, England, 14, pp 319-328.

12. Finney, D. J., and G. C. Varley (1955). An Example of the Truncated Poisson Distribution, Biometrics, 11, pp 387-394.

BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
BEGIN LAST LINE OF TEXT
3/8"

EPA-2B7 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

13. Fisher, R. A. (1958). Statistical Methods for Research Workers, Thirteenth Revised Edition, Hafner Publishing Company, Incorporated, New York, New York, 356 pp.

14. Fisher, R. A., A. S. Corbet, and C. B. Williams (1943). The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population, Journal of Animal Ecology, 12, pp 42-57.

15. Hartley, H. O. (1961). The Modified Gauss-Newton Method for Fitting of Non-Linear Regression Functions by Least Squares, Technometrics, 3,2, pp 269-280.

16. Haseman, J. K., and E. R. Soares (1976). The Distribution of Fetal Deaths in Control Mice and Its Implications on Statistical Tests for Dominant Lethal Effects, Mutation Research, 41, pp 277-288.

17. Hasselblad, V. (1969). Estimation of Finite Mixtures of Distributions from the Exponential Family, Journal of the American Statistical Association, 64, pp 1459-1571.

18. Irwin, J. O. (1959). On the Estimation of the Mean of a Poisson Distribution with the Zero Class Missing, Biometrics, 15, pp 324-326.

19. Johnson, N. I., and S. Kotz (1969). Discrete Distributions, John Wiley and Sons, Incorporated, Somerset, New Jersey, 328 pp.

20. Kendall, M. G., and A. Stuart (1961). The Advanced Theory of Statistics, Volume 2, Hafner Publishing Company, New York, New York, 676 pp.

21. Love, G. J., S. Lan, C. M. Shy, and R. J. Struba (1980). The Incidence and Severity of Acute Respiratory Illness in Families Exposed to Different Levels of Air Pollution, New York Metropolitan Area, 1971-1972, Dept. of Epidemiology, School of Public Health, Univ. of N. C., Chapel Hill, North Carolina.

22. Mantel, N. (1951). Evaluation of a Class of Diagnostic Tests, Biometrics, 7, pp 240-246.

23. McKendrick, A. G. (1926). Applications of Mathematics to Medical Problems, Proceedings of Edinburgh Mathematical Society, 44, pp 98-130.

24. Montmort, P. R. (1714). Essai D'Analyse Sur Les Jeux de Hasards, Paris, France.

25. Morgan, C., W. Nelson, and P. Caporal (1979). General Maximum Likelihood Fitting with STATPAC, General Electric Technical Report.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin.) (4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD. BEGIN SECTIONS HERE

26. Nelson, W. C., and H. A. David (1967). The Logarithmic Distribution: A Review, Virginia Journal of Science, 18, pp 95-102.

27. Nie, N. H., C. H. Hull, J. G. Jenkins, K. Steinbrenner, and D. H. Brent (1975). Statistical Package for the Social Sciences, Second Edition, McGraw Hill Book Company, New York, New York, 673 pp.

28. Orchard, T., and M. A. Woodbury (1972). A Missing Information Principle: Theory and Applications, Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability, 1, pp 697-715.

29. Pascal, B. (1679). Varia Opera Mathematica D. Petri de Fermat (Tolossae).

30. Potthoff, R. F., and M. Whittinghill (1966). Testing for Homogeneity, I. The binomial and multinomial distributions, Biometrika, 53, pp 167-182.

31. Potthoff, R. F., and M. Whittinghill (1966). Testing for Homogeneity, II. The Poisson distribution, Biometrika, 53, pp 183-190.

32. Rider, P. R. (1961). Estimating the Parameters of Mixed Poisson, Binomial, and Weibull Distributions by the Method of Moments, Bulletin of International Statistical Institute, Vol. 39, pp 225-232.

33. Sampford, M. R. (1955). The Truncated Negative Binomial Distribution, Biometrika, 42, pp 58-69.

34. Schilling, W. (1947). A Frequency Distribution Represented on the Sum of Two Poisson Distributions, Journal of American Statistical Association, 42, pp 407-424.

35. Skellam, J. G. (1948). A Probability Distribution Derived from the Binomial Distribution by Regarding the Probability of Success as Variable Between the Sets of Trials, Journal of the Royal Statistical Society, Series B, London, 10, pp 257-261.

36. Sokal, R. R. and F. J. Rohlf (1973). Introduction to Biostatistics, Freeman and Company, San Francisco, California, pp 66-67.

37. Wyshak, G. (1974). A Program for Estimating the Parameters of the Truncated Negative Binomial Distribution, Journal of Applied Statistics, 23, 1, pp 87-91.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA. OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

APPENDIX A.

INDEX

Ames test data, 22, 28
Bernoulli, James, 8
Beta-binomial distribution, 17
Beta distribution, 3, 17
Binomial distribution, 1, 8, 15, 17, 30
Chromosome break data, 35, 38
Constant distribution, 5
Cumulative frequency distribution, 5
Dice data, Weldon's, 9
Digamma function, 18
Discrete rectangular distribution, 5
Discrete uniform distribution, 5
EM algorithm, 11, 14, 24, 27
Exponential distribution, 1
Frequency distribution, 5
Gall-fly eggs, 12, 25
Gamma distribution, 1
Gauss-Newton method, 3, 17
Geometric distribution, 1, 30
Goodness of fit, 1, 6, 9, 12, 15, 19, 21, 25, 28, 32, 35, 38
Homogeneity test 9, 22
Illness data, 32
Likelihood ratio test, 3, 6, 9, 15, 25, 32
Logarithmic distribution, 35, 37
Lognormal distribution, 1
Maximum likelihood, 3
Mendelian factors, 6
Mice data, 19
Missing information principle, 3, 11, 14, 24, 27
Mixture of two binomials, 14
Mixture of two Poissons, 27
Moment estimates, 14, 27, 30
Moments, sample, 5
Negative binomial distribution, 3, 30, 34
Newton-Raphson method, 3, 31, 34, 37
Normal distribution, 1
Pascal distribution, 30, 32
Pearson curves, 1

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

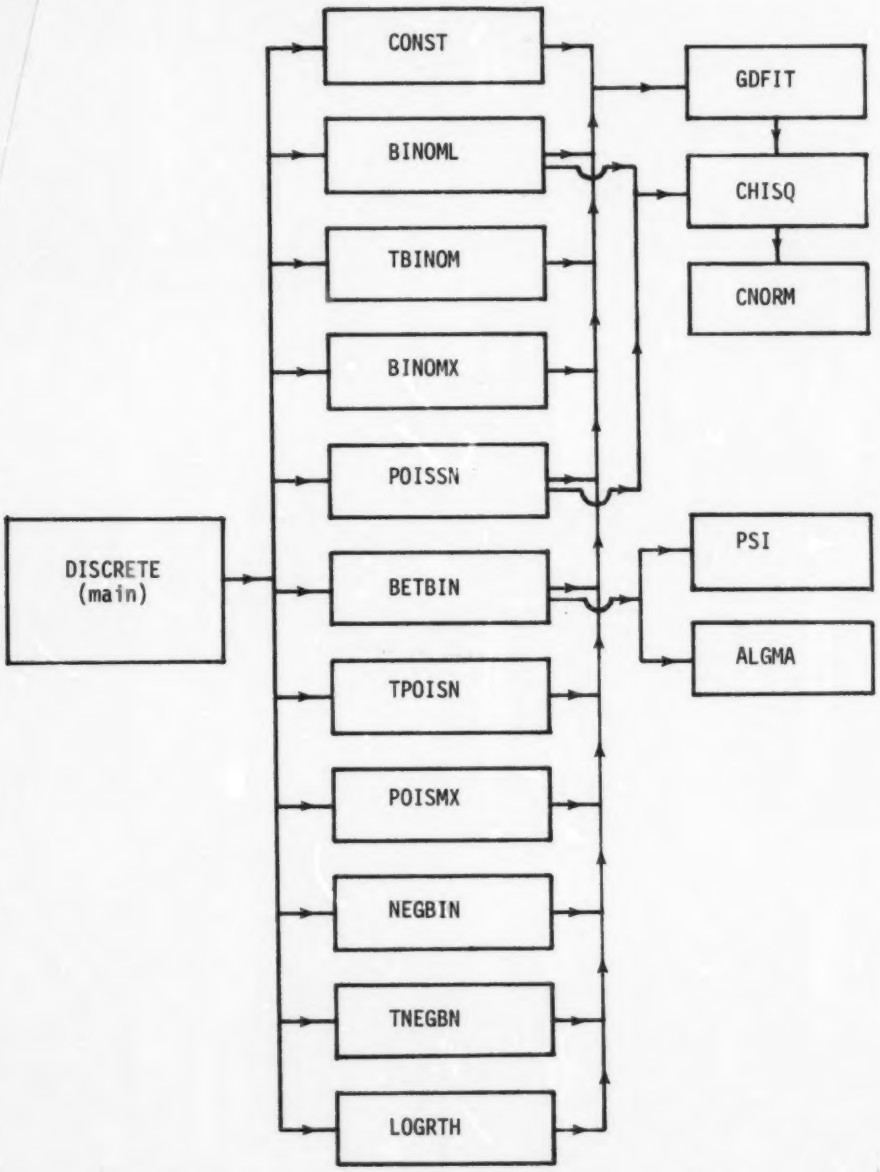
EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE
6-1/2"
9-1/8"
BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"
PAGE NUMBER
EPA-287 (Cin.)
(4-76)

Poisson distribution, 21, 28
Positive binomial distribution, 11
Positive Poisson distribution, 24
Primula counts, 6
Rectangular distribution, discrete, 5
Sample moments, 5
Sex distribution data, 15
Singularities, 3
Sokal and Rohlf's data, 15
Truncated binomial distribution, 11
Truncated negative binomial 34, 37
Truncated Poisson distribution, 12, 24
Uniform distribution, 1
Uniform distribution, discrete, 5
Variance test, 22
Varley's data, 12, 25
Weibul, 1
Weldon's dice data, 9

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD, BEGIN SECTIONS HERE

APPENDIX B. FLOW CHART OF THE DISCRETE PROGRAM



BEGIN LAST LINE OF TEXT
3/8"
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

APPENDIX C. ERROR MESSAGES

INVALID COMMAND:

The name of the distribution fitted does not match any names in the program.

INVALID PARAMETER SPECIFICATION:

The values for one or more of the parameters (XP's) are outside the allowable limits. These limits are given in the definition section of each distribution.

PROCEDURE DID NOT CONVERGE:

This message can occur in any of the routines which require iterative methods to obtain the estimates. There is no easy way to correct the problem, although the estimates given by the program still may be useful. We would appreciate receiving such examples so that we can improve our procedures in the future.

INSUFFICIENT EXPECTED FREQUENCIES:

The goodness-of-fit routine (GDFIT) did not find sufficient expected frequencies to perform a goodness-of-fit test.

UNEXPECTED END OF FILE ENCOUNTERED:

An end-of-file was encountered while reading a card other than a title card. Check sequence of input cards.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

APPENDIX D.

LISTING OF THE PROGRAMS

```
C     DISCRETE DISTRIBUTION FITTING PACKAGE              REVISED  9-24-80
C          PROGRAMMERS: ANDERSON, HASSELBLAD
C
C     MAIN PROGRAM
C
C     THE FOLLOWING ARE VARIABLES IN COMMON:
C     NF IS AN ARRAY OF OBSERVED COUNTS
C     E IS AN ARRAY OF EXPECTED COUNTS
C     XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C     K IS THE VALUE CORRESPONDING TO THE LARGEST OBSERVED FREQUENCY.
C     M IS THE LENGTH OF THE E ARRAY
C     INT IS AN OPTIONAL INPUT INTEGER.
C     NP IS TH NUMBER OF PARAMETERS EXTIMATED
C     IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C        IF = 2 FOR TRUNCATED DISTRIBUTIONS.
C     NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C
      DIMENSION FMT(20)
      COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
      CHARACTER*8 NC,NCON(12)
      DATA NCON/'CONSTANT','BINOMIAL','TBINOMAL','BINOMIX ','BETBINOM',
     1 'POISSON ','TPOISSON','POISSMIX','NEGBINOM','TNEGBINM','LOGARITH'
     2 ,'        '/
      DATA ND/12/
 2    CONTINUE
      NTOT = 0
      DO 3 I = 1, 200
      NF(I) = 0
 3    CONTINUE
C
C     NC IS THE NAME OF THE DISTRIBUTION TO BE FITTED
C     READ IN THE CONTROL CARDS
C
      READ (5,5,END=910) TITLE
 5    FORMAT(20A4)
      READ (5,10,END=920) K
 10   FORMAT(I5)
      IF(K.GE.200) GO TO 930
      KP = K + 1
      READ (5,5,END=920) FMT
      READ (5,FMT,END=920)  (NF(I),I=1,KP)
      DO 4 I = 1, KP
         NTOT = NTOT + NF(I)
 4    CONTINUE
 6    CONTINUE
      READ (5,15,END=920) NC, INT, XP
 15   FORMAT(A8,I7,4F10.0)
      DO 7 I = 1, ND
         IF(NC.NE.NCON(I)) GO TO 7
         GO TO (11,12,13,14,16,17,18,19,21,22,23,2) I
 7    CONTINUE
      WRITE(6,55) NC, K,XP
 55   FORMAT('0',10X,'INVALID COMMAND: ',A8,I5,4F10.4)
      GO TO 6
 11   CONTINUE
      CALL CONST
```

```
      GO TO 6
12    CONTINUE
      CALL BINOML
      GO TO 6
13    CONTINUE
      CALL TBINOM
      GO TO 6
14    CONTINUE
      CALL BINOMX
      GO TO 6
16    CONTINUE
      CALL BETBIN
      GO TO 6
17    CONTINUE
      CALL POISSN
      GO TO 6
18    CONTINUE
      CALL TPOISN
      GO TO 6
19    CONTINUE
      CALL POISMX
      GO TO 6
21    CONTINUE
      CALL NEGBIN
      GO TO 6
22    CONTINUE
      CALL TNEGBN
      GO TO 6
23    CONTINUE
      CALL LOGRTH
      GO TO 6
910   CONTINUE
      STOP
920   CONTINUE
      WRITE (6,925)
925   FORMAT('1',10X,'UNEXPECTED END-OF-FILE ENCOUNTERED')
      STOP
930   CONTINUE
      WRITE (6,935) K
935   FORMAT('1',10X,'VALUE OF K TOO LARGE:',I7)
      STOP
      END
```

```
C     CNORM                                                REVISED  5- 9-80
C        PROGRAMMER: HASSELBLAD
C     COMPUTES THE NORMAL PROBABILITY INTEGRAL FROM MINUS INFINITY TO X
C     REFERENCE: APPROX. FOR DIG. COMP. - HASTINGS, P. 169
      FUNCTION CNORM(X)
      DATA S2P,P,A1,A2,A3,A4,A5,S2/1.12837917 ,.3275911,.225836846,
     1  -.252128668,1.25969513,-1.28782245 ,.94064607,.707106781/
      XN = 1./(1. + P*ABS(X*S2))
      PHI = S2P*EXP(-X**2/2.)
      PHIS=1.-XN*(A1+XN*(A2+XN*(A3+XN*(A4+XN*A5))))*PHI
      TERM = .5*PHIS
      IF(X.LT.0.) TERM = -TERM
      CNORM = 0.5 + TERM
      RETURN
      END
```

```
C      CONSTANT DISTRIBUTION FITTING ROUTINE INCLUDING SAMPLE MOMENTS
C           PROGRAMMERS: HASSELBLAD, ANDERSON          REVISED  5- 5-80
C
       SUBROUTINE CONST
       COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
C
C      NF IS AN ARRAY OF OBSERVED COUNTS
C      E IS AN ARRAY OF EXPECTED COUNTS
C      XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C      K IS THE VALUE CORRESPONDING TO THE LARGEST OBSERVED FREQUENCY.
C      M IS THE LENGTH OF THE E ARRAY
C      INT IS AN OPTIONAL INPUT INTEGER.
C      NP IS THE NUMBER OF PARAMETERS ESTIMATED.
C      IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C         AND IF = 2 FOR TRUNCATED DISTRIBUTIONS.
C      NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C      P IS THE ESTIMATED PARAMETER OF THE BINOMIAL DISTRIBUTION: B(N,P)
C
       WRITE (6,10) TITLE
 10    FORMAT('1',10X,20A4)
       M = K + 1
       TOT = NTOT
       NFL = 0
       PSUM = 0.
       DO 4 I = 1, 4
       E(I) = XP(I)
       IF(E(I).GT.0.) NFL = 1
 4     CONTINUE
       IF(NFL.EQ.0) GO TO 6
       IF(M.LE.4) GO TO 7
       READ (5,5) (E(I),I=5,M)
 5     FORMAT(8F10.0)
       GO TO 7
 6     CONTINUE
C
C      ASSUME A UNIFORM DISTRIBUTION IF NOT SPECIFIED.
C
       DO 11 I = 1, M
          E(I) = 1.
 11    CONTINUE
 7     CONTINUE
C
C      SUM UP EXPECTED FREQUENCIES SO THEY CAN BE NORMALIZED.
C
       DO 12 I = 1, M
          IF(E(I).LT.0.) GO TO 900
          PSUM = PSUM + E(I)
 12    CONTINUE
       XLK = 0.
C
C      NORMALIZE EXPECTED FREQUENCIES AND COMPUTE LOG-LIKELIHOOD.
C
       DO 8 I = 1, M
          E(I) = E(I)/PSUM
          IF(E(I).GT.0.) XLK = XLK + NF(I)*ALOG(E(I))
          E(I) = E(I)*TOT
 8     CONTINUE
       IF = 1
       IF(E(1).EQ.0) IF = 2
       XM1 = 0.
       XM2 = 0.
       XM3 = 0.
       XM4 = 0.
       NP = INT
C
C      COMPUTE SAMPLE MOMENTS
C
       DO 2 I = 1, M
          X = I - 1
          XM1 = XM1 + NF(I)*X
```

```
2     CONTINUE
      XM1 = XM1/TOT
      DO 3 I = 1, M
         X = I - 1 - XM1
         X2 = X*X
         X3 = X2*X
         X4 = X2*X2
         XM2 = XM2 + NF(I)*X2
         XM3 = XM3 + NF(I)*X3
         XM4 = XM4 + NF(I)*X4
3     CONTINUE
      XM2 = XM2/TOT
      XM3 = XM3/TOT
      XM4 = XM4/TOT
C
C     WRITE OUT RESULTS
C
      WRITE (6,15) XM1
15    FORMAT(1H0,10X,'SAMPLE MEAN =',5X,F14.6)
      WRITE (6,20) XM2
20     FORMAT(1H ,10X,'SAMPLE VARIANCE =',F15.6)
      WRITE (6,25) XM3
25    FORMAT(1H ,10X,'THIRD MOMENT =',4X,F14.6)
      WRITE (6,30) XM4
30    FORMAT(1H ,10X,'FOURTH MOMENT =',3X,F14.6)
      WRITE (6,40)
40    FORMAT('0',10X,'VALUE  COUNT  PERCENT CUMULATIVE CUMULATIVE')
      WRITE (6,45)
45    FORMAT(' ',37X,'COUNT',4X,'PERCENT')
      NCUM = 0
      CUMP = 0.
      DO 9 I = 1, M
         J = I - 1
         PCT = NF(I)*100/TOT
         NCUM = NCUM + NF(I)
         CUMP = CUMP + PCT
         IF(E(I).GT.0..OR.NF(I).GT.0) WRITE (6,50) J,NF(I),PCT,NCUM,CUMP
50       FORMAT(' ',10X,I5,I7,F9.2,I11,F11.2)
9     CONTINUE
C
C     COMPUTE GOODNESS OF FIT
C
      CALL GDFIT
      WRITE (6,35) XLK
35    FORMAT('0',10X,'LOG-LIKELIHOOD =',F12.4)
      RETURN
900   CONTINUE
      WRITE (6,905) INT, (XP(I),I=1,4)
905   FORMAT('0',10X,'INVALID PARAMETER SPECIFICATION'/11X,I7,4F11.5)
      IF(M.GT.4) WRITE (6,910) (E(I),I=5,M)
910   FORMAT(' ',10X,8F11.5)
      RETURN
      END
```

```
C      SUBROUTINE TO FIT A BINOMIAL DISTRIBUTION          REVISED  9-25-80
C         PROGRAMMERS: HASSELBLAD, ANDERSON
C
       SUBROUTINE BINOML
       COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
C
C      NF IS AN ARRAY OF OBSERVED COUNTS
C      E IS AN ARRAY OF EXPECTED COUNTS
C      XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C      K IS THE VALUE CORRESPONDING TO THE LARGEST OBSERVED FREQUENCY.
C      M IS THE LENGTH OF THE E ARRAY
C      INT IS AN OPTIONAL INPUT INTEGER.
C      NP IS THE NUMBER OF PARAMETERS ESTIMATED
C      IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C         AND IF = 2 FOR TRUNCATED DISTRIBUTIONS.
C      NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C      P IS THE ESTIMATED PARAMETER OF THE BINOMIAL DISTRIBUTION: B(N,P)
C
       IF = 1
       WRITE (6,30) TITLE
 30    FORMAT('1',10X,20A4)
       WRITE (6,5)
 5     FORMAT ('0',10X,'FIT TO A BINOMIAL DISTRIBUTION')
       WRITE (6,40)
 40    FORMAT('0',10X,'P(X,N) = N!P**X(1-P)**(N-X)/(X!(N-X)!)')
C
C      CHECK FOR IMPROPER PARAMETER VALUES.
C
       IF(XP(1).LT.0.OR.XP(1).GT.1.) GO TO 900
       IF(INT.LT.0) GO TO 900
       N = MAX(INT,K)
       CN = N
       KP = K + 1
       M = MIN(KP+1,N+1)
       TOT = NTOT
       NP = 1
       IF(XP(1).GT.0.) NP = 0
       P = XP(1)
C
C      COMPUTE MOMENTS
C
       XSUM = 0.
       YSUM = 0.
       X2 = 0.
       DO 2 I = 1, KP
          CI = I - 1
          XSUM = XSUM + CI*NF(I)
          YSUM = YSUM + (CN - CI)*NF(I)
          X2 = X2 + CI**2*NF(I)
 2     CONTINUE
C
C      ESTIMATE P IF NOT SPECIFIED
C
       IF(XP(1).EQ.0.) P = XSUM/(TOT*CN)
       WRITE (6,10) NTOT, N
 10    FORMAT ('0',10X,'TOTAL SAMPLE SIZE =',I6,', N =',I4)
       IF(XP(1).EQ.0.) WRITE (6,15) P
 15    FORMAT('0',10X,'ESTIMATED VALUE OF P =',F6.4)
       IF(XP(1).GT.0.) WRITE (6,20) P
 20    FORMAT('0',10X,'SPECIFIED VALUE OF P =',F6.4)
       VAR = P*(1. - P)/(TOT*CN)
       IF(XP(1).EQ.0.) WRITE (6,35) VAR
 35    FORMAT(' ',10X,'VARIANCE OF P =',E13.5)
C
C      COMPUTE EXPECTED COUNTS, LOG-LIKELIHOOD
C
       XLK = 0.
       Q = 1. - P
       ALP = ALOG(P)
       ALQ = ALOG(Q)
```

```
      XDC = 1.
      XNC = CN
      TERM = CN*ALQ
      DO 3 I = 1, M
         E(I) = EXP(TERM)*TOT
         XLK = XLK + NF(I)*TERM
         TERM = TERM + ALP - ALQ + ALOG(XNC) - ALOG(XDC)
         IF(XNC.GT.1.) XNC = XNC - 1.
         XDC = XDC + 1.
 3    CONTINUE
C
C     COMPUTE GOODNESS OF FIT TESTS
C
      NDF = NTOT
      IF(XP(1).EQ.0.) NDF = NTOT - 1
      VART = X2/(P*Q*CN) - CN*TOT*P/Q
      PVAL = CHISQ(VART,NDF)
      WRITE (6,45) VART, PVAL
 45   FORMAT('0',10X,'HOMOGENEITY TEST =',F9.2,', P =',F6.4)
      CALL GDFIT
      WRITE (6,50) XLK
 50   FORMAT ('0',10X,'LOG-LIKELIHOOD  =',F12.4)
      RETURN
 900  CONTINUE
      WRITE (6,905) INT,XP(1)
 905  FORMAT ('0',10X,'INVALID PARAMETER SPECIFICATION'/11X,I7,F11.5)
      RETURN
      END
```

```
C     ALGMA                                                REVISED  5- 9-80
C        PROGRAMMER: HASSELBLAD
C     COMPUTES LOG(GAMMA(X)) USING STIRLING'S APPROXIMATION
C
      REAL FUNCTION ALGMA(X)
      Z = X
      SUM = 0.
      IF(X.LE.0.) GO TO 9
 3    CONTINUE
      IF(Z.GE.5.) GO TO 2
      SUM = SUM - ALOG(Z)
      Z = Z + 1
      GO TO 3
 2    CONTINUE
      TERM = 1. + .083333333/Z + .0034722222/Z**2 - .0026813272/Z**3
      ALGMA = (Z - .5)*ALOG(Z) - Z + .91893853 + ALOG(TERM) + SUM
      RETURN
 9    CONTINUE
      ALGMA = -0.
      RETURN
      END
```

```
C     SUBROUTINE TO FIT A TRUNCATED BINOMIAL DISTRIBUTION
C        PROGRAMMERS: HASSELBLAD, ANDERSON              REVISED 10-16-80
C
      SUBROUTINE TBINOM
      COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
C
C     NF IS AN ARRAY OF OBSERVED COUNTS
C     E IS AN ARRAY OF EXPECTED COUNTS
C     XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C     K IS THE VALUE CORRESPONDING TO THE LARGEST OBSERVED FREQUENCY
C     M IS THE LENGTH OF THE E ARRAY
C     INT IS AN OPTIONAL INPUT INTEGER.
C     NP IS THE NUMBER OF PARAMETERS ESTIMATED
C     IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C        AND IF = 2 FOR TRUNCATED DISTRIBUTIONS.
C     NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C     PT IS THE NEW ESTIMATE OF THE PARAMETER P
C     P IS THE OLD ESTIMATE OF THE PARAMETER P
C     ZN IS THE ESTIMATE OF THE UNOBSERVED CELL F(0)
C
      WRITE (6,30) TITLE
 30   FORMAT('1',10X,20A4)
      WRITE (6,5)
 5    FORMAT('0',10X,'FIT TO A TRUNCATED BINOMIAL DISTRIBUTION')
C
C     CHECK FOR IMPROPER PARAMETER VALUES.
C
      IF(XP(1).LT.0.0.OR.XP(1).GT.1.0) GO TO 900
      IF (INT.LT.0) GO TO 900
      KP = K + 1
      NP = 1
      IF(XP(1).GT.0.) NP = 0
      N = MAX(INT,K)
      M = MIN(KP+1,N+1)
      IF = 2
      CN = N
      NT = 0
      P = XP(1)
      XM1 = 0.
      DO 2 I = 2, KP
         CI = I - 1
         NT = NT + NF(I)
         XM1 = XM1 + CI*NF(I)
 2    CONTINUE
      TOT = NT
      XM1 = XM1/TOT
C
C     COMPUTE THE INITIAL ESTIMATE OF P IF NOT SPECIFIED.
C
      IF(XP(1).GT.0.) GO TO 6
      P = (XM1 - NF(2)/TOT)/(CN - NF(2)/TOT)
      WRITE (6,10) NT, N
 10   FORMAT('0',10X,'TOTAL SAMPLE SIZE =',I6,', N =',I4)
C
C     BEGIN ITERATION PROCEDURE
C
      DO 4 IT = 1, 50
         QN = (1. - P)**N
         ZN = QN*TOT/(1. - QN)
         PT = 0.
         DO 3 I = 2, KP
            CI = I - 1.
            PT = PT + CI*NF(I)/((TOT + ZN)*CN)
 3       CONTINUE
         IF(ABS(PT-P).LE.0.00001) GO TO 7
         P = PT
 4    CONTINUE
      WRITE (6,95)
 95   FORMAT('0',10X,'PROCEDURE DID NOT CONVERGE')
 7    CONTINUE
```

```
      P = PT
 6    CONTINUE
      IF(XP(1).EQ.0.) WRITE (6,40) P
 40   FORMAT('0',10X,'ESTIMATED VALUE OF P =',F6.4)
      IF(XP(1).GT.0.) WRITE (6,45) P
 45   FORMAT('0',10X,'SPECIFIED VALUE OF P =',F6.4)
      XL = 0.
      PT = P
      QT = 1. - PT
      ALP = ALOG(PT)
      ALQ = ALOG(QT)
      XDC = 2.
      XNC = CN - 1.
      CUM = 0.
      QN = 1. - QT**N
      TERM = ALOG(CN) + ALP + XNC*ALQ - ALOG(QN)
C
C     COMPUTE EXPECTED COUNTS, ASYMPTOTIC VARIANCE
C
      E(1) = 0.
      VAR = -TOT*(CN*(CN - 1.)*QT**(N - 2)/QN + (QT**(N - 1)*CN)**2/QN)
      DO 8 I = 2, M
         CI = I - 1.
         QI = QT**(I - 1)
         VAR = VAR + CI*NF(I)/PT**2 + (CN - CI)*NF(I)/QT**2
         E(I) = EXP(TERM)*TOT
         CUM = CUM + E(I)
         XL = XL + NF(I)*TERM
         IF(I.NE.M) TERM = TERM + ALP - ALQ + ALOG(XNC) - ALOG(XDC)
         IF(XNC.GT.1.) XNC = XNC - 1.
         XDC = XDC + 1.
 8    CONTINUE
      VAR = 1./VAR
      IF(XP(1).EQ.0.) WRITE (6,50) VAR
 50   FORMAT(' ',10X,'ASYMP. VAR. OF P =',E13.5)
C
C     COMPUTE GOODNESS OF FIT TESTS.
C
      CALL GDFIT
      WRITE (6,20) XL
 20   FORMAT('0',10X,'LOG-LIKELIHOOD  =',F12.4)
      RETURN
 900  CONTINUE
      WRITE (6,905) INT, XP(1)
 905  FORMAT('0',10X,'INVALID PARAMETER SPECIFICATION'/11X,I7,F11.5)
      RETURN
      END
```

```
C     SUBROUTINE TO FIT A BETA BINOMIAL DISTRIBUTION    REVISED 4-21-80
C          PROGRAMMER: HASSELBLAD
C
      SUBROUTINE BETBIN
      COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
C
C     NF IS AN ARRAY OF OBSERVED COUNTS
C     E IS AN ARRAY OF EXPECTED COUNTS
C     XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C     M IS THE LENGTH OF THE E ARRAY
C     K IS THE VALUE CORRESPONDING TO THE LARGEST OBSERVED FREQUENCY
C     INT IS AN OPTIONAL INPUT INTEGER.
C     NP IS THE NUMBER OF PARAMETERS ESTIMATED
C     IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C        AND IF = 2 FOR TRUNCATED DISTRIBUTIONS.
C     NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C     ALPHA AND BETA ARE THE PARAMETERS TO BE ESTIMATED (GIVEN AS  A
C          AND  B  IN THE FORMULA OF FORMAT 10)
C
      DATA EPS/0.00001/
      WRITE (6,30) TITLE
 30   FORMAT('1',10X,20A4)
C
C     CHECK FOR IMPROPER PARAMETER VALUES
C
      IF(XP(1).LT.0..OR.XP(2).LT.0.) GO TO 900
      IF = 1
      N = MAX(INT,K)
      CN = N
      KP = K + 1
      M = MIN(KP+1,N+1)
      TOT = NTOT
      NP = 2
      IF(XP(1).GT.0.) NP = NP - 1
      IF(XP(2).GT.0.) NP = NP - 1
      WRITE (6,5)
 5    FORMAT('0',10X,'FIT TO A BETA-BINOMIAL DISTRIBUTION')
      WRITE (6,10)
 10   FORMAT('0',10X,'P(X) = N!(A+X-1)!(B+N-X-1)!(A-1)!(B-1)!/'/18X,
     1 '(X!(N-X)!(A+B+N-1)!(A+B-1)!')
      WRITE (6,15) NTOT, N
 15   FORMAT('0',10X,'TOTAL SAMPLE SIZE =',I6,', N =',I4)
C
C     COMPUTE FACTORIAL MOMENTS FOR INITIAL ESTIMATES
C
      R1 = 0.
      R2 = 0.
      DO 6 I = 1, KP
         CI = I - 1
         R1 = R1 + CI*NF(I)/TOT
         R2 = R2 + CI*(CI - 1.)*NF(I)/TOT
 6    CONTINUE
      R2 = R2/R1
C
C     ESTIMATE THE INITIAL VALUES FOR ALPHA AND BETA
C
      ALPHA = (R1*R2 - (CN - 1.)*R1)/((CN - 1.)*R1 - CN*R2)
      IF(XP(1).GT.0.) ALPHA = XP(1)
      BETA = ALPHA*(CN/R1 - 1.)
      IF(XP(2).GT.0.) BETA = XP(2)
      IF(XP(1).GT.0..AND.XP(2).GT.0.) GO TO 16
C
C     BEGIN GAUSS-NEWTON ITERATION SCHEME
C
      DO 12 IT = 1, 100
         ALBE = ALPHA + BETA
         DELTA = 1.
C
C        COMPUTE CURRENT LIKELIHOOD FUNCTION
C
```

```
C         COMPUTE ELEMENTS OF FIRST PARTIAL DERIVATIVES
C         SUM THEIR CROSS PRODUCTS
C
          XLK1 = TOT*(ALGMA(CN + 1.) + ALGMA(ALBE) - ALGMA(ALPHA)
     1     - ALGMA(BETA) - ALGMA(ALBE + CN))
          T = PSI(ALBE) - PSI(ALBE + CN)
          T1 = T - PSI(ALPHA)
          T2 = T - PSI(BETA)
          D1 = 0.
          D2 = 0.
          D11 = 0.
          D12 = 0.
          D22 = 0.
          DO 2 I = 1, KP
             CI = I - 1
             XLK1 = XLK1 + NF(I)*(ALGMA(ALPHA + CI) + ALGMA(BETA + CN-CI)
     1        - ALGMA(CI + 1.) - ALGMA(CN - CI + 1.))
             E1 = T1 + PSI(ALPHA + CI)
             E2 = T2 + PSI(BETA + CN - CI)
             D1 = D1 + E1*NF(I)
             D2 = D2 + E2*NF(I)
             D11 = D11 + E1**2*NF(I)
             D12 = D12 + E1*E2*NF(I)
             D22 = D22 + E2**2*NF(I)
 2        CONTINUE
C
C         CHECK FOR SPECIFIED PARAMETERS
C
          IF(XP(1).GT.0..OR.XP(2).GT.0.) D12 = 0.
          IF(XP(1).GT.0.) D1 = 0.
          IF(XP(1).GT.0.) D11 = 1.
          IF(XP(2).GT.0.) D2 = 0.
          IF(XP(2).GT.0.) D22 = 1.
          DET = D22*D11 - D12*D12
          TEMP = D11
          D11 = D22/DET
          D22 = TEMP/DET
          D12 = -D12/DET
          DELA = D11*D1 + D12*D2
          DELB = D22*D2 + D12*D1
C
C         COMPUTE CONSTRAINED NEW ESTIMATES AND THEIR LIKELIHOOD
C         THE ESTIMATES ARE CONSTRAINED TO BE GREATER THAN EPS
C
          ALPHP = AMAX1(ALPHA + DELA,EPS)
          BETAP = AMAX1(BETA + DELB,EPS)
          ALBEP = ALPHP + BETAP
          XLK2 = TOT*(ALGMA(CN + 1.) + ALGMA(ALBEP) - ALGMA(ALPHP)
     1     - ALGMA(BETAP) - ALGMA(ALBEP + CN))
          DO 3 I = 1, KP
             CI = I - 1
             XLK2 = XLK2 + NF(I)*(ALGMA(ALPHP + CI) + ALGMA(BETAP+CN-CI)
     1        - ALGMA(CI + 1.) - ALGMA(CN - CI + 1.))
 3        CONTINUE
C
C         FIND BEST ESTIMATES USING A CHOPPING METHOD
C
 7        CONTINUE
          DELTA = DELTA*0.5
          XLK3 = XLK2
C
C         IF CHOPPING GETS TOO SMALL, QUIT
C
          IF(DELTA.LE.0.000001.AND.IT.GT.1) GO TO 16
          IF(DELTA.LE.0.000001) GO TO 18
          ALPHP = AMAX1(ALPHA + DELTA*DELA,EPS)
          BETAP = AMAX1(BETA + DELTA*DELB,EPS)
          ALBEP = ALPHP + BETAP
          XLK2 = TOT*(ALGMA(CN + 1.) + ALGMA(ALBEP) - ALGMA(ALPHP)
     1     - ALGMA(BETAP) - ALGMA(ALBEP + CN))
```

```
      DO 4 I = 1, KP
         CI = I - 1
         XLK2 = XLK2 + NF(I)*(ALGMA(ALPHP + CI) + ALGMA(BETAP+CN-CI)
     1     - ALGMA(CI + 1.) - ALGMA(CN - CI + 1.))
 4    CONTINUE
      IF(XLK2.LE.XLK1) GO TO 7
C
C     OPTIMIZE STEP LENGTH
C
      DOPT = (XLK3 - 4.*XLK2 + 3.*XLK1)/(2.*(XLK3 - 2.*XLK2 + XLK1))
      IF(DOPT.GE.1..OR.DOPT.LE.0.) DOPT = 1.
      DELTA = DELTA*DOPT
C
C     SAVE NEW IMPROVED ESTIMATES
C
      ALPHA = AMAX1(ALPHA + DELTA*DELA,EPS)
      BETA = AMAX1(BETA + DELTA*DELB,EPS)
C
C     CHECK FOR CONVERGENCE - A CHANGE IN PARAMETERS OF LESS
C     THAN EPS
C
      IF(DELTA*(ABS(DELA) + ABS(DELB)).LE.EPS) GO TO 16
 12   CONTINUE
 18   CONTINUE
      WRITE (6,95)
 95   FORMAT('0PROCEDURE DID NOT CONVERGE')
 16   CONTINUE
C
C     COMPUTE FINAL LIKELIHOOD, EXPECTED COUNTS
C
      XLK = 0.
      TERM = ALGMA(BETA + CN) - ALGMA(ALPHA + BETA + CN) +
     1 ALGMA(ALPHA + BETA) - ALGMA(BETA)
      DO 17 I = 1, M
         CI = I - 1
         E(I) = TOT*EXP(TERM)
         XLK = XLK + NF(I)*TERM
         TERM = TERM + ALOG(ALPHA + CI) - ALOG(CI + 1.)
         IF(I.NE.M ) TERM = TERM + ALOG(CN - CI) - ALOG(BETA+CN-CI-1.)
 17   CONTINUE
      IF(XP(2).GT.0.) D22 = 0.
      IF(XP(1).GT.0.) D11 = 0.
C
C     WRITE OUT ESTIMATES AND THEIR VARIANCES
C
      IF(XP(1).GT.0.) WRITE (6,20) ALPHA
 20   FORMAT('0',10X,'SPECIFIED VALUE OF ALPHA =',F12.5)
      IF(XP(1).EQ.0.) WRITE (6,25) ALPHA
 25   FORMAT('0',10X,'ESTIMATED VALUE OF ALPHA =',F12.5)
      IF(XP(2).GT.0.) WRITE (6,35) BETA
 35   FORMAT(' ',10X,'SPECIFIED VALUE OF BETA  =',F12.5)
      IF(XP(2).EQ.0.) WRITE (6,40) BETA
 40   FORMAT(' ',10X,'ESTIMATED VALUE OF BETA  =',F12.5)
      IF(XP(1).GT.0..AND.XP(2).GT.0.) GO TO 19
      WRITE (6,45)
 45   FORMAT('0',10X,'ASYMPTOTIC COVARIANCE MATRIX OF ESTIMATES')
      WRITE (6,50)
 50   FORMAT(' ',21X,'ALPHA',8X,'BETA')
      WRITE (6,55) D11, D12, D12, D22
 55   FORMAT(' ',10X,'ALPHA',2E13.5/11X,'BETA ',2E13.5)
 19   CONTINUE
C
C     COMPUTE GOODNESS OF FIT
C
      CALL GDFIT
      WRITE (6,65) XLK
 65   FORMAT('0',10X,'LOG-LIKELIHOOD =',F12.4)
      RETURN
 900  CONTINUE
      WRITE (6,905) INT, XP(1), XP(2)
```

```
905   FORMAT('0',10X,'INVALID PARAMETER SPECIFICATION'/11X,I7,4F11.5)
      RETURN
      END
```

```
C     CHISQ                                                REVISED  5- 9-80
C        PROGRAMMER: HASSELBLAD
C     COMPUTES THE CHI-SQUARE INTEGRAL TO X FOR N DEGREES OF FREEDOM
C        REFERENCE: HANDBOOK OF MATHEMATICAL FUNCTIONS
C
      REAL FUNCTION CHISQ(X,N)
      SUM = 0.
      IF(X.GT.0.0.AND.N.GT.0)GO TO 2
      CHISQ = 1.
      RETURN
 2    CONTINUE
      IF(N.GT.50) GO TO 6
      IF((N/2)*2.NE.N)GO TO 11
      IF(N.LE.2)GO TO 3
      M = N/2 - 1
      TERM = 1.
      DO 4 I = 1, M
      CI = 2.*FLOAT(I)
      TERM = TERM*X/CI
      SUM = SUM + TERM
 4    CONTINUE
 3    CONTINUE
      CHISQ = EXP(-X/2.0)* 1.0+SUM)
      RETURN
 11   CONTINUE
      TERM = 1.0/X
      IF(N.LE.1)GO TO 12
      M = (N-1)/2
      DO 13 I = 1, M
      CI = 2.*FLOAT(I)-1.
      TERM = TERM*X/CI
      SUM = SUM + TERM
 13   CONTINUE
 12   CONTINUE
      XS = SQRT(X)
      CHISQ = 2. - 2.*CNORM(XS) + XS*EXP(-X/2.)*.7978846*SUM
      RETURN
 6    CONTINUE
      CHISQ = 1.-CNORM(((X/N)**.3333333-1.+2./(9.*N))/SQRT(2./(9.*N)))
      RETURN
      END
```

```
C     SUBROUTINE TO FIT A MIXTURE OF TWO BINOMIAL DISTRIBUTIONS
C          PROGRAMMER: HASSELBLAD                          REVISED  9-25-80
C
      SUBROUTINE BINOMX
      COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
C
C     NF IS AN ARRAY OF OBSERVED COUNTS
C     E IS AN ARRAY OF EXPECTED COUNTS
C     XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C     K IS THE VALUE CORRESPONDING TO THE LARGEST OBSERVED FREQUENCY
C     M IS THE LENGTH OF THE E ARRAY
C     INT IS AN OPTIONAL INPUT INTEGER.
C     NP IS THE NUMBER OF PARAMETERS ESTIMATED
C     IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C        AND IF = 2 FOR TRUNCATED DISTRIBUTIONS.
C     NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C     P1 IS THE ESTIMATED PARAMETER OF THE FIRST BINOMIAL DISTRIBUTION
C     P2 IS THE ESTIMATED PARAMETER OF THE SECOND BINOMIAL DISTRIBUTION
C     ALPHA IS THE ESTIMATED PROPORTION CORRESPONDING TO THE FIRST
C        BINOMIAL DISTRIBUTION
C
      IF = 1
      WRITE (6,20) TITLE
 20   FORMAT('1',10X,20A4)
C
C     CHECK FOR IMPROPER PARAMETER VALUES
C
      IF(XP(1).LT.0..OR.XP(1).GT.1.) GO TO 900
      IF(XP(2).LT.0..OR.XP(2).GT.1.) GO TO 900
      IF(XP(3).LT.0..OR.XP(3).GT.1.) GO TO 900
      N = MAX(INT,K)
      CN = N
      KP = K + 1
      M = MIN(KP+1,N+1)
      NP = 3
      IF(XP(1).GT.0.) NP = NP - 1
      IF(XP(2).GT.0.) NP = NP - 1
      IF(XP(3).GT.0.) NP = NP - 1
      P = 0.
C
C     COMPUTE ESTIMATE OF AVERAGE P
C
      DO 2 I = 1, KP
         P = P + (I-1)*NF(I)
 2    CONTINUE
      TOT = NTOT
      P = P/(TOT*CN)
      WRITE (6,5)
 5    FORMAT ('0',10X,'FIT TO A MIXTURE OF TWO BINOMIAL DISTRIBUTIONS')
      WRITE (6,15)
 15   FORMAT('0',10X,'P(X) = N!(ALPHA(P1**X)(1-P1)**(N-X) +'/15X,
     1 '(1-ALPHA)(P2**X)(1-P2)**(N-X))/(X!(N-X)!)')
      WRITE (6,10) NTOT, N
 10   FORMAT ('0',10X,'TOTAL SAMPLE SIZE =',I6,', N =',I4)
C
C     COMPUTE INITIAL ESTIMATES
C
      ALPHA = 0.5
      IF(XP(3).GT.0.) ALPHA = XP(3)
      P1 = 0.75*P
      IF(P1.LE.0.05) P1 = 0.05
      IF(XP(1).GT.0.) P1 = XP(1)
      P2 = 1.333*P
      IF(P2.GE.0.95) P2 = 0.95
      IF(XP(2).GT.0.) P2 = XP(2)
      IF(XP(1).GT.0..AND.XP(2).GT.0..AND.XP(3).GT.0.) GO TO 8
C
C     BEGIN MIP-EM ITERATION SCHEME
C
      DO 6 IT = 1, 1000
```

```
      X1 = 0.
      X2 = 0.
      XS = 0.
      Q1 = 1. - P1
      Q2 = 1. - P2
      ALP1 = ALOG(P1)
      ALQ1 = ALOG(Q1)
      ALP2 = ALOG(P2)
      ALQ2 = ALOG(Q2)
      XDC = 1.
      XNC = CN
      TERM1 = CN*ALQ1
      TERM2 = CN*ALQ2
      DO 7 I = 1, KP
         CI = I - 1
         D1 = EXP(TERM1)
         D2 = EXP(TERM2)
         D = ALPHA*D1 + (1. - ALPHA)*D2
         X1 = X1 + D1*CI*NF(I)/D
         X2 = X2 + D2*CI*NF(I)/D
         XS = XS + ALPHA*D1*NF(I)/D
         IF(I.NE.M) TERM1 = TERM1 + ALP1-ALQ1 + ALOG(XNC) - ALOG(XDC)
         IF(I.NE.M) TERM2 = TERM2 + ALP2-ALQ2 + ALOG(XNC) - ALOG(XDC)
         IF(XNC.GT.1.) XNC = XNC - 1.
         XDC = XDC + 1.
 7    CONTINUE
      X1 = X1/(TOT*CN)
      X2 = X2/(TOT*CN)
      XS = XS/TOT
      IF(XP(1).GT.0.) X1 = XP(1)
      IF(XP(2).GT.0.) X2 = XP(2)
      IF(XP(3).GT.0.) XS = XP(3)
C
C     TEST FOR CONVERGENCE
C
      IF(ABS(P1-X1) + ABS(P2-X2) + ABS(ALPHA-XS).LE.0.0001) GO TO 4
      P1 = X1
      P2 = X2
      ALPHA = XS
 6  CONTINUE
    WRITE (6,95)
 95 FORMAT('0',10X,'PROCEDURE DID NOT CONVERGE')
 4  CONTINUE
    P1 = X1
    P2 = X2
    ALPHA = XS
 8  CONTINUE
C
C   COMPUTE FINAL LIKELIHOOD, EXPECTED FREQUENCIES,
C   AND SECOND PARTIAL DERIVATIVES
C
    XLK = 0.
    Q1 = 1. - P1
    Q2 = 1. - P2
    ALP1 = ALOG(P1)
    ALQ1 = ALOG(Q1)
    ALP2 = ALOG(P2)
    ALQ2 = ALOG(Q2)
    XDC = 1.
    XNC = CN
    TERM1 = CN*ALQ1
    TERM2 = CN*ALQ2
    ESUM = 0.
    D11 = 0.
    D12 = 0.
    D22 = 0.
    D13 = 0.
    D23 = 0.
    D33 = 0.
    DO 9 I = 1, KP
```

```
          CI = I - 1
          D1 = EXP(TERM1)
          D2 = EXP(TERM2)
          D = ALPHA*D1 + (1. - ALPHA)*D2
          E(I) = TOT*D
          ESUM = ESUM + E(I)
          XLK = XLK + NF(I)*ALOG(D)
          IF(I.NE.M) TERM1 = TERM1 + ALP1 - ALQ1 + ALOG(XNC) - ALOG(XDC)
          IF(I.NE.M) TERM2 = TERM2 + ALP2 - ALQ2 + ALOG(XNC) - ALOG(XDC)
          IF(XNC.GT.1.) XNC = XNC - 1.
          XDC = XDC + 1.
          T1 = NF(I)/D
          T2 = -T1/D
          T3 = D1*ALPHA
          T4 = D2*(1. - ALPHA)
          T5 = CI/P1 - (CN - CI)/(1. - P1)
          T6 = CI/P2 - (CN - CI)/(1. - P2)
          D11 = D11 + T2*T3**2*T5**2 + T1*T3*(T5**2 - CI/P1**2 -
     1     (CN - CI)/(1. - P1)**2)
          D22 = D22 + T2*T4**2*T6**2 + T1*T4*(T6**2 - CI/P2**2 -
     1     (CN - CI)/(1. - P2)**2)
          D12 = D12 + T2*T3*T4*T5*T6
          D13 = D13 + T2*T3*T5*(D1 - D2)
          D23 = D23 + T2*T4*T6*(D1 - D2)
          D33 = D33 + T2*(D1 - D2)**2
 9     CONTINUE
       E(KP) = E(KP) + TOT - ESUM
C
C      ALLOW FOR SPECIFIED PARAMETERS
C
       IF(XP(1).GT.0.) D11 = 1.
       IF(XP(2).GT.0.) D22 = 1.
       IF(XP(3).GT.0.) D33 = 1.
       IF(XP(1).GT.0..OR.XP(2).GT.0.) D12 = 0.
       IF(XP(1).GT.0..OR.XP(3).GT.0.) D13 = 0.
       IF(XP(2).GT.0..OR.XP(3).GT.0.) D23 = 0.
       DET = D11*D22*D33 + 2.*D12*D23*D13 - D13**2*D22 - D23**2*D11 -
     1 D12**2*D33
C
C      WRITE OUT ESTIMATES AND THEIR VARIANCES
C
       IF(XP(1).GT.0.) WRITE (6,25) P1
 25    FORMAT('0',10X,'SPECIFIED VALUE OF P1 =',F11.5)
       IF(XP(1).EQ.0.) WRITE (6,30) P1
 30    FORMAT('0',10X,'ESTIMATED VALUE OF P1 =',F11.5)
       IF(XP(2).GT.0.) WRITE (6,35) P2
 35    FORMAT(' ',10X,'SPECIFIED VALUE OF P2 =',F11.5)
       IF(XP(2).EQ.0.) WRITE (6,40) P2
 40    FORMAT(' ',10X,'ESTIMATED VALUE OF P2 =',F11.5)
       IF(XP(3).GT.0.) WRITE (6,45) ALPHA
 45    FORMAT(' ',10X,'SPECIFIED VALUE OF ALPHA =',F8.5)
       IF(XP(3).EQ.0.) WRITE (6,50) ALPHA
 50    FORMAT(' ',10X,'ESTIMATED VALUE OF ALPHA =',F8.5)
       IF(XP(1).GT.0..AND.XP(2).GT.0..AND.XP(3).GT.0.) GO TO 11
       V11 = (D23**2 - D22*D33)/DET
       V12 = (D12*D33 - D13*D23)/DET
       V13 = (D13*D22 - D12*D23)/DET
       V22 = (D13**2 - D11*D33)/DET
       V23 = (D11*D23 - D13*D12)/DET
       V33 = (D12**2 - D11*D22)/DET
       IF(XP(1).GT.0.) V11 = 0.
       IF(XP(2).GT.0.) V22 = 0.
       IF(XP(3).GT.0.) V33 = 0.
       WRITE (6,55)
 55    FORMAT('0',10X,'ASYMPTOTIC COVARIANCE MATRIX OF ESTIMATES')
       WRITE (6,60)
 60    FORMAT(' ',21X,'  P1 ',8X,'  P2 ',8X,'ALPHA')
       WRITE (6,65) V11, V12, V13, V12, V22, V23, V13, V23, V33
 65    FORMAT(' ',10X,'P1    ',3E13.5/11X,'P2    ',3E13.5/11X,'ALPHA',
     1 3E13.5)
```

```
 11   CONTINUE
C
C     COMPUTE GOODNESS OF FIT
C
      CALL GDFIT
      WRITE (6,70) XLK
 70   FORMAT('0',10X,'LOG-LIKELIHOOD =',F12.4)
      RETURN
 900  CONTINUE
      WRITE (6,905) INT, XP(1), XP(2), XP(3)
 905  FORMAT('0',10X,'INVALID PARAMETER SPECIFICATION'/11X,I7,4F11.5)
      RETURN
      END
```

```
C     SUBROUTINE TO FIT A POISSON DISTRIBUTION          REVISED  9-25-80
C        PROGRAMMERS: HASSELBLAD, ANDERSON
      SUBROUTINE POISSN
      COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
C
C     NF IS AN ARRAY OF OBSERVED COUNTS
C     E IS AN ARRAY OF EXPECTED COUNTS
C     XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C     K IS THE VALUE CORRESPONDING TO THE LARGEST OBSERVED FREQUENCY
C     M IS THE LENGTH OF THE E ARRAY
C     INT IS AN OPTIONAL INPUT INTEGER.
C     NP IS THE NUMBER OF PARAMETERS ESTIMATED
C     IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C        AND IF = 2 FOR TRUNCATED DISTRIBUTIONS.
C     NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C     XLMB IS THE ESTIMATE OF THE PARAMETER LAMBDA
C
      WRITE (6,30) TITLE
 30   FORMAT ('1',10X,20A4)
      WRITE (6,5)
 5    FORMAT ('0',10X,'FIT TO A POISSON DISTRIBUTION')
      WRITE (6,40)
 40   FORMAT('0',10X,'P(X) = EXP(-LAMBDA)LAMBDA**X/X!')
      IF(XP(1).LT.0.) GO TO 900
      IF = 1
      KP = K + 1
      M = K + 2
      NP = 1
      IF(XP(1).GT.0) NP = 0
      TOT = NTOT
      XLMB = XP(1)
      IF(XP(1).GT.0.) GO TO 4
C
C     COMPUTE ESTIMATE OF LAMBDA
C
      XM2 = 0.
      DO 2 I = 1, KP
         CI = I - 1
         XLMB = XLMB + CI*NF(I)
         XM2 = XM2 + CI**2*NF(I)
 2    CONTINUE
      XLMB = XLMB/TOT
 4    CONTINUE
```

```
      WRITE (6,10) NTOT
 10   FORMAT ('0',10X,'TOTAL SAMPLE SIZE =',I6)
      IF(XP(1).EQ.0.) WRITE (6,15) XLMB
 15   FORMAT('0',10X,'ESTIMATED VALUE OF LAMBDA =',F9.4)
      IF(XP(1).GT.0.) WRITE (6,20) XLMB
 20   FORMAT('0',10X,'SPECIFIED VALUE OF LAMBDA =',F9.4)
      VAR = XLMB/TOT
      IF(XP(1).EQ.0.) WRITE (6,25) VAR
 25   FORMAT(' ',10X,'VARIANCE OF LAMBDA =',E13.5)
      XLK = 0.
      ALP = ALOG(XLMB)
      CUM = 0.
      XDC = 1.
      TERM = -XLMB
C
C     COMPUTE EXPECTED COUNTS, LOG-LIKELIHOOD
C
      DO 3 I = 1, KP
         E(I) = EXP(TERM)*TOT
         CUM = CUM + E(I)
         XLK = XLK + NF(I)*TERM
         TERM = TERM + ALP - ALOG(XDC)
         XDC = XDC + 1.
 3    CONTINUE
      E(M) = TOT - CUM
C
C     COMPUTE GOODNESS OF FIT TESTS
C
      NDF = NTOT - 1
      VART = XM2/XLMB - TOT*XLMB
      PVAL = CHISQ(VART,NDF)
      IF(XP(1).EQ.0.) WRITE (6,45) VART, PVAL
 45   FORMAT('0',10X,'VARIANCE TEST =',F8.2,', P =',F6.4)
      CALL GDFIT
      WRITE (6,35) XLK
 35   FORMAT('0',10X,'LOG-LIKELIHOOD  =',F12.4)
      RETURN
 900  CONTINUE
      WRITE (6,905) XP(1)
 905  FORMAT('0',10X,'INVALID PARAMETER SPECIFIED'/11X,F11.5)
      RETURN
      END
```

```
C     SUBROUTINE TO FIT A TRUNCATED POISSON DISTRIBUTION
C        PROGRAMMERS: HASSELBLAD, ANDERSON              REVISED  9-25-80
C
      SUBROUTINE TPOISN
      COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
C
C     NF IS AN ARRAY OF OBSERVED COUNTS
C     E IS AN ARRAY OF EXPECTED COUNTS
C     XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C     K IS THE VALUE CORRESPONDING TO THE LARGEST OBSERVED FREQUENCY
C     M IS THE LENGTH OF THE E ARRAY
C     INT IS AN OPTIONAL INPUT INTEGER.
C     NP IS THE NUMBER OF PARAMETERS ESTIMATED
C     IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C        AND IF = 2 FOR TRUNCATED DISTRIBUTIONS.
C     NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C     XLMB IS THE CURRENT ESTIMATE OF THE PARAMETER LAMBDA
C     XLMBP IS THE NEW ESTIMATE OF THE PARAMETER LAMBDA
C     ZN IS THE ESTIMATE OF THE UNOBSERVED CELL F(0)
C
      WRITE (6,30) TITLE
 30   FORMAT('1',10X,20A4)
      WRITE (6,5)
 5    FORMAT('0',10X,'FIT TO A TRUNCATED POISSON DISTRIBUTION')
      WRITE (6,20)
 20   FORMAT('0',10X,'P(X) = EXP(-LAMBDA)LAMBDA**X/(X!(1-EXP(-LAMBDA)))'
     1)
      KP = K + 1
      M = K + 2
      IF = 2
      NT = 0
      NP = 1
      IF(XP(1).GT.0.) NP = 0
      IF(XP(1).LT.0.) GO TO 900
      XM1 = 0.
      XM2 = 0.
C
C     COMPUTE ESTIMATE OF LAMBDA AND F(0)
C
      XLMB = XP(1)
      DO 2 I = 2, KP
         CI = I - 1
         NT = NT + NF(I)
         XM1 = XM1 + CI*NF(I)
         XM2 = XM2 + CI*(CI - 1.)*NF(I)
 2    CONTINUE
      TOT = NT
      ZN = XM1**2/XM2 - TOT
      WRITE (6,10) NT
 10   FORMAT('0',10X,'TOTAL SAMPLE SIZE =',I6)
      IF(XP(1).GT.0.) GO TO 6
      XLMB = XM1/(ZN + TOT)
C
C     BEGIN ITERATION PROCEDURE
C
      DO 4 IT = 1, 50
         EXLMB = EXP(-XLMB)
         ZN = EXLMB*TOT/(1. - EXLMB)
         XLMBP = XM1/(ZN + TOT)
         IF(ABS(XLMBP - XLMB)/AMAX1(XLMB,1.).LE.0.00001) GO TO 7
         XLMB = XLMBP
 4    CONTINUE
      WRITE (6,40)
 40   FORMAT('0',10X,'PROCEDURE DID NOT CONVERGE')
 7    CONTINUE
      XLMB = XLMBP
 6    CONTINUE
      IF(XP(1).EQ.0.) WRITE (6,15) XLMB
 15   FORMAT('0',10X,'ESTIMATED VALUE OF LAMBDA =',F8.4)
      IF(XP(1).GT.0.) WRITE (6,25) XLMB
```

```
 25   FORMAT('0',10X,'SPECIFIED VALUE OF LAMBDA =',F8.4)
C
CC     COMPUTE ESTIMATED VARIANCE OF ESTIMATE
C
       VAR = XLMB**2/(XM1*(XLMB + 1. - XM1/TOT))
       IF(XP(1).EQ.0)WRITE (6,45) VAR
 45    FORMAT(' ',10X,'ASYMP. VAR. OF LAMBDA =',E13.5)
       XL = 0.
       ALP = ALOG(XLMB)
       CUM = 0.
       XDC = 2.
       TERM = ALP - XLMB - ALOG(1. - EXP(-XLMB))
C
CC     COMPUTE EXPECTED COUNTS
C
       E(1) = 0.
       DO 3 I = 2, KP
          E(I) = EXP(TERM)*TOT
          CUM = CUM + E(I)
          XL = XL + NF(I)*TERM
          TERM = TERM + ALP - ALOG(XDC)
          XDC = XDC + 1.
  3    CONTINUE
       E(M) = TOT - CUM
C
CC     COMPUTE GOODNESS OF FIT
C
       CALL GDFIT
       WRITE (6,60) XL
 60    FORMAT('0',10X,'LOG-LIKELIHOOD =',F12.4)
       RETURN
 900   CONTINUE
       WRITE (6,905) XP(1)
 905   FORMAT('0',10X,'INVALID PARAMETER SPECIFICATION'/11X,F11.5)
       RETURN
       END
```

```
C      PSI                                                    REVISED  7-14-80
CC        PROGRAMMER: HASSELBLAD
CC     COMPUTES THE PSI FUNCTION (ALSO KNOWN AS THE DIGAMMA FUNCTION):
CC     PSI(Z) = D(LN(GAMMA(Z)))/DZ.
CC
CC     THE PROGRAM USES AN ASYMPTOTIC FORMULA AS GIVEN ON P. 259 OF
CC     "HANDBOOK OF MATHEMATICAL FUNCTIONS".
C
       REAL FUNCTION PSI(Z)
       X = Z
       S = 0.
  2    CONTINUE
       IF(X.GE.5.) GO TO 3
       S = S + 1./X
       X = X + 1.
       GO TO 2
  3    CONTINUE
       PSI = ALOG(X) - .5/X - (.083333333/(X + .1/(X + .0031349208/X)))
      1 /X - S
       RETURN
       END
```

```
C      SUBROUTINE TO FIT A MIXTURE OF TWO POISSON DISTRIBUTIONS
C           PROGRAMMER: HASSELBLAD                          REVISED  4-30-80
C
       SUBROUTINE POISMX
       COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
C
C      NF IS AN ARRAY OF OBSERVED COUNTS
C      E IS AN ARRAY OF EXPECTED COUNTS
C      XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C      K IS THE VALUE CORRESPONDING TO THE LARGEST OBSERVED FREQUENCY
C      M IS THE LENGTH OF THE E ARRAY
C      INT IS AN OPTIONAL INPUT INTEGER.
C      NP IS THE NUMBER OF PARAMETERS ESTIMATED
C      IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C         AND IF = 2 FOR TRUNCATED DISTRIBUTIONS.
C      NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C      XLM1 IS THE ESTIMATE OF THE PARAMETER LAMBDA1
C      XLM2 IS THE ESTIMATE OF THE PARAMETER LAMBDA2
C      ALPHA IS THE ESTIMATED PROPORTION CORRESPONDING TO THE FIRST
C         POISSON DISTRIBUTION
C
       IF = 1
       WRITE (6,20) TITLE
 20    FORMAT('1',10X,20A4)
C
C      CHECK FOR IMPROPER PARAMETER VALUES
C
       IF(XP(1).LT.0.) GO TO 900
       IF(XP(2).LT.0.) GO TO 900
       IF(XP(3).LT.0..OR.XP(3).GT.1.) GO TO 900
       KP = K + 1
       M = K + 2
       NP = 3
       IF(XP(1).GT.0.) NP = NP - 1
       IF(XP(2).GT.0.) NP = NP - 1
       IF(XP(3).GT.0.) NP = NP - 1
       XLM = 0.
C
C      COMPUTE ESTIMATE OF AVERAGE P
C
       DO 2 I = 1, KP
          XLM = XLM + (I-1)*NF(I)
 2     CONTINUE
       TOT = NTOT
       XLM = XLM/TOT
       WRITE (6,5)
 5     FORMAT ('0',10X,'FIT TO A MIXTURE OF TWO POISSON DISTRIBUTIONS')
       WRITE (6,15)
 15    FORMAT('0',10X,'P(X) = ALPHA*EXP(-LAMBDA1)X**LAMBDA1/X! +'/18X,
      1 '(1-ALPHA)EXP(-LAMBDA2)X**LAMBDA2)/X!')
       WRITE (6,10) NTOT
 10    FORMAT ('0',10X,'TOTAL SAMPLE SIZE =',I6)
C
C      COMPUTE INITIAL ESTIMATES
C
       ALPHA = 0.5
       IF(XP(3).GT.0.) ALPHA = XP(3)
       XLM1 = 0.75*XLM
       IF(XP(1).GT.0.) XLM1 = XP(1)
       XLM2 = 1.3333*XLM
       IF(XP(2).GT.0.) XLM2 = XP(2)
       IF(XP(1).GT.0..AND.XP(2).GT.0..AND.XP(3).GT.0.) GO TO 8
C
C      BEGIN MIP-EM ITERATION SCHEME
C
       DO 6 IT = 1, 2000
          X1 = 0.
          X2 = 0.
          XS = 0.
          ALM1 = ALOG(XLM1)
```

```
      ALM2 = ALOG(XLM2)
      XDC = 1.
      TERM1 = -XLM1
      TERM2 = -XLM2
      DO 7 I = 1, KP
         CI = I - 1
         D1 = EXP(TERM1)
         D2 = EXP(TERM2)
         D = ALPHA*D1 + (1. - ALPHA)*D2
         X1 = X1 + D1*CI*NF(I)/D
         X2 = X2 + D2*CI*NF(I)/D
         XS = XS + ALPHA*D1*NF(I)/D
         TERM1 = TERM1 + ALM1 - ALOG(XDC)
         TERM2 = TERM2 + ALM2 - ALOG(XDC)
         XDC = XDC + 1.
7     CONTINUE
      X1 = X1/TOT
      X2 = X2/TOT
      XS = XS/TOT
      IF(XP(1).GT.0.) X1 = XP(1)
      IF(XP(2).GT.0.) X2 = XP(2)
      IF(XP(3).GT.0.) XS = XP(3)
C
C     TEST FOR CONVERGENCE
C
      IF(ABS(XLM1-X1)+ABS(XLM2-X2)+ABS(ALPHA-XS).LE.0.0001) GO TO 4
      XLM1 = X1
      XLM2 = X2
      ALPHA = XS
6     CONTINUE
      WRITE (6,95)
95    FORMAT('0',10X,'PROCEDURE DID NOT CONVERGE')
4     CONTINUE
      XLM1 = X1
      XLM2 = X2
      ALPHA = XS
8     CONTINUE
C
C     COMPUTE FINAL LIKELIHOOD, EXPECTED FREQUENCIES,
C     AND SECOND PARTIAL DERIVATIVES
C
      XLK = 0.
      ALM1 = ALOG(XLM1)
      ALM2 = ALOG(XLM2)
      XDC = 1.
      TERM1 = -XLM1
      TERM2 = -XLM2
      ESUM = 0.
      D11 = 0.
      D12 = 0.
      D22 = 0.
      D13 = 0.
      D23 = 0.
      D33 = 0.
      DO 9 I = 1, KP
         CI = I - 1
         D1 = EXP(TERM1)
         D2 = EXP(TERM2)
         D = ALPHA*D1 + (1. - ALPHA)*D2
         E(I) = TOT*D
         ESUM = ESUM + E(I)
         XLK = XLK + NF(I)*ALOG(D)
         TERM1 = TERM1 + ALM1 - ALOG(XDC)
         TERM2 = TERM2 + ALM2 - ALOG(XDC)
         XDC = XDC + 1.
         T1 = NF(I)/D
         T2 = -T1/D
         T3 = D1*ALPHA
         T4 = D2*(1. - ALPHA)
         T5 = (CI/XLM1 - 1.)
```

```
          T6 = (CI/XLM2 - 1.)
          D11 = D11 + T2*T3**2*T5**2 + T1*T3*(T5**2 - CI/XLM1**2)
          D22 = D22 + T2*T4**2*T6**2 + T1*T4*(T6**2 - CI/XLM2**2)
          D12 = D12 + T2*T3*T4*T5*T6
          D13 = D13 + T2*T3*T5*(D1 - D2)
          D23 = D23 + T2*T4*T6*(D1 - D2)
          D33 = D33 + T2*(D1 - D2)**2
    9     CONTINUE
          E(M) = TOT - ESUM
C
C         ALLOW FOR SPECIFIED PARAMETERS
C
          IF(XP(1).GT.0.) D11 = 1.
          IF(XP(2).GT.0.) D22 = 1.
          IF(XP(3).GT.0.) D33 = 1.
          IF(XP(1).GT.0..OR.XP(2).GT.0.) D12 = 0.
          IF(XP(1).GT.0..OR.XP(3).GT.0.) D13 = 0.
          IF(XP(2).GT.0..OR.XP(3).GT.0.) D23 = 0.
          DET = D11*D22*D33 + 2.*D12*D23*D13 - D13**2*D22 - D23**2*D11 -
         1 D12**2*D33
C
C         WRITE OUT ESTIMATES AND THEIR VARIANCES
C
          IF(XP(1).GT.0.) WRITE (6,25) XLM1
   25     FORMAT('0',10X,'SPECIFIED VALUE OF LAMBDA1 =',F11.5)
          IF(XP(1).EQ.0.) WRITE (6,30) XLM1
   30     FORMAT('0',10X,'ESTIMATED VALUE OF LAMBDA1 =',F11.5)
          IF(XP(2).GT.0.) WRITE (6,35) XLM2
   35     FORMAT(' ',10X,'SPECIFIED VALUE OF LAMBDA2 =',F11.5)
          IF(XP(2).EQ.0.) WRITE (6,40) XLM2
   40     FORMAT(' ',10X,'ESTIMATED VALUE OF LAMBDA2 =',F11.5)
          IF(XP(3).GT.0.) WRITE (6,45) ALPHA
   45     FORMAT(' ',10X,'SPECIFIED VALUE OF ALPHA =',F13.5)
          IF(XP(3).EQ.0.) WRITE (6,50) ALPHA
   50     FORMAT(' ',10X,'ESTIMATED VALUE OF ALPHA =',F13.5)
          IF(XP(1).GT.0..AND.XP(2).GT.0..AND.XP(3).GT.0.) GO TO 11
          V11 = (D23**2 - D22*D33)/DET
          V12 = (D12*D33 - D13*D23)/DET
          V13 = (D13*D22 - D12*D23)/DET
          V22 = (D13**2 - D11*D33)/DET
          V23 = (D11*D23 - D13*D12)/DET
          V33 = (D12**2 - D11*D22)/DET
          IF(XP(1).GT.0.) V11 = 0.
          IF(XP(2).GT.0.) V22 = 0.
          IF(XP(3).GT.0.) V33 = 0.
          WRITE (6,55)
   55     FORMAT('0',10X,'ASYMPTOTIC COVARIANCE MATRIX OF ESTIMATES')
          WRITE (6,60)
   60     FORMAT(' ',22X,'LAMBDA1',6X,'LAMBDA2',6X,'ALPHA')
          WRITE (6,65) V11, V12, V13, V12, V22, V23, V13, V23, V33
   65     FORMAT(' ',10X,'LAMBDA1',3E13.5/11X,'LAMBDA2',3E13.5/11X,'ALPHA',
         1 2X,3E13.5)
   11     CONTINUE
C
C         COMPUTE GOODNESS OF FIT
C
          CALL GDFIT
          WRITE (6,70) XLK
   70     FORMAT('0',10X,'LOG-LIKELIHOOD =',F12.4)
          RETURN
   900    CONTINUE
          WRITE (6,905) INT, XP(1), XP(2), XP(3)
   905    FORMAT('0',10X,'INVALID PARAMETER SPECIFICATION'/11X,I7,4F11.5)
          RETURN
          END
```

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD. BEGIN SECTIONS HERE

```
C     SUBROUTINE TO FIT A NEGATIVE BINOMIAL DISTRIBUTION
C        PROGRAMMER: HASSELBLAD                          REVISED  9-30-80
C
      SUBROUTINE NEGBIN
      COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
C
C     NF IS AN ARRAY OF OBSERVED COUNTS
C     E IS AN ARRAY OF EXPECTED COUNTS
C     XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C     K IS THE VALUE CORRESPONDING TO THE LARGEST OBSERVED FREQUENCY
C     M IS THE LENGTH OF THE E ARRAY
C     INT IS AN OPTIONAL INPUT INTEGER.
C     NP IS THE NUMBER OF PARAMETERS ESTIMATED
C     IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C        AND IF = 2 FOR TRUNCATED DISTRIBUTIONS.
C     NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C     P IS THE ESTIMATE OF THE PARAMETER P
C     XN IS THE ESTIMATE OF THE PARAMETER N
C     XNP IS THE IMPROVED ESTIMATE OF THE PARAMETER N
C     N1, N2, CN1, CN2, XLK1, XLK2, TERM1, TERM2, ALT1, ALT2  ARE USED
C        TO FIND THE OPTIMAL INTEGER SOLUTION FOR N
C
      WRITE (6,30) TITLE
 30   FORMAT ('1',10X,20A4)
      WRITE (6,5)
 5    FORMAT('0',10X,'FIT TO A NEGATIVE BINOMIAL DISTRIBUTION')
      WRITE (6,10)
 10   FORMAT('0',10X,'P(X) = (N+X-1)!P**X/(X!(N-1)!(1+P)**(N+X))')
      WRITE (6,15) NTOT
 15   FORMAT('0',10X,'TOTAL SAMPLE SIZE =',I6)
      IF(XP(1).LT.0..OR.XP(2).LT.0.) GO TO 900
      IF = 1
      KP = K + 1
      M = K + 2
      NP = 2
      IF(XP(1).GT.0.) NP = NP - 1
      IF(XP(2).GT.0.) NP = NP - 1
      TOT = NTOT
C
C     COMPUTE MOMENTS
C
      XM1 = 0.
      XM2 = 0.
      DO 2 I = 1, KP
         CI = I - 1
         XM1 = XM1 + CI*NF(I)
         XM2 = XM2 + CI**2*NF(I)
 2    CONTINUE
      XM1 = XM1/TOT
      V = XM2/TOT - XM1**2
C
C     COMPUTE INITIAL ESTIMATES
C
      P = XP(1)
      XN = XP(2)
      IF(XP(1).GT.0..AND.XP(2).GT.0.) GO TO 4
      IF(V.LE.XM1) GO TO 920
      IF(XP(1).EQ.0..AND.XP(2).EQ.0.) P = V/XM1 - 1.
      IF(XP(1).EQ.0..AND.XP(2).GT.0.) P = XM1/XP(2)
      IF(XP(2).EQ.0.) XN = XM1/P
      IF(XP(2).GT.0.) GO TO 4
C
C     USE NEWTON-RAPHSON ITERATION SCHEME FOR N IF NOT SPECIFIED
C
      DO 7 IT = 1, 100
         SUM = 0.
         SUM1 = 0.
         S = 0.
         S1 = 0.
         TERM = XN - 1.
```

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA. OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cim.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

```
            DO 6 I = 1, K
               TERM = TERM + 1.
               S = S + 1./TERM
               S1 = S1 + 1./TERM**2
               SUM = SUM + S*NF(I+1)
               SUM1 = SUM1 + S1*NF(I+1)
6           CONTINUE
            IF(XP(1).EQ.0.) F = TOT*ALOG(1. + XM1/XN) - SUM
            IF(XP(1).GT.0.) F = TOT*ALOG(1. + P) - SUM
            IF(XP(1).EQ.0.) F1 = -(TOT*XM1)/(XN**2 + XM1*XN) + SUM1
            IF(XP(1).GT.0.) F1 = SUM1
            XNP = XN - F/F1
            IF(XNP.LE.0.) GO TO 9
            IF(ABS((XNP - XN)/XN).LE.0.0001) GO TO 8
            XN = XNP
7       CONTINUE
9       CONTINUE
        WRITE (6,95)
95      FORMAT('0',10X,'PROCEDURE DID NOT CONVERGE')
8       CONTINUE
        XN = XNP
        IF(XP(1).EQ.0.) P = XM1/XN
4       CONTINUE
        NXB = XM1
        N1 = XN
        N2 = N1 + 1
        IF(N1.LE.0)N1 = 1
        CN1 = N1
        CN2 = N2
        P1 = XP(1)
        P2 = XP(1)
        IF(XP(1).EQ.0.) P1 = XM1/CN1
        IF(XP(1).EQ.0.) P2 = XM1/CN2
C
C       COMPUTE LIKELIHOOD, EXPECTED COUNTS
C
        XLK = 0.
        XLK1 = 0.
        XLK2 = 0.
        ESUM = 0.
        XNC = XN
        XNC1 = CN1
        XNC2 = CN2
        Q = 1. + P
        Q1 = 1. + P1
        Q2 = 1. + P2
        ALT = ALOG(P) - ALOG(Q)
        ALT1 = ALOG(P1) - ALOG(Q1)
        ALT2 = ALOG(P2) - ALOG(Q2)
        TERM = -XN*ALOG(Q)
        TERM1 = -CN1*ALOG(Q1)
        TERM2 = -CN2*ALOG(Q2)
        DO 17 I = 1, KP
           CI = I - 1
           E(I) = TOT*EXP(TERM)
           ESUM = ESUM + E(I)
           XLK = XLK + NF(I)*TERM
           XLK1 = XLK1 + NF(I)*TERM1
           XLK2 = XLK2 + NF(I)*TERM2
           TERM = TERM + ALT + ALOG(XNC) - ALOG(CI + 1.)
           TERM1 = TERM1 + ALT1 + ALOG(XNC1) - ALOG(CI + 1.)
           TERM2 = TERM2 + ALT2 + ALOG(XNC2) - ALOG(CI + 1.)
           XNC = XNC + 1.
           XNC1 = XNC1 + 1.
           XNC2 = XNC2 + 1.
17      CONTINUE
        E(M) = TOT - ESUM
        D11 = -TOT*(XN*P**2 - 2.*P*XM1 - XM1)/(P*Q)**2
        D12 = TOT/Q
        D22 = SUM1
```

BEGIN LAST LINE OF TEXT HERE
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD. BEGIN SECTIONS HERE

```
C
C     WRITE OUT ESTIMATES AND THEIR VARIANCES
C
      IF(XP(1).GT.0.) WRITE (6,20) P
 20   FORMAT('0',10X,'SPECIFIED VALUE OF P =',F12.5)
      IF(XP(1).EQ.0.) WRITE (6,25) P
 25   FORMAT('0',10X,'ESTIMATED VALUE OF P =',F12.5)
      IF(XP(2).GT.0.) WRITE (6,35) XN
 35   FORMAT(' ',10X,'SPECIFIED VALUE OF N =',F12.5)
      IF(XP(2).EQ.0.) WRITE (6,40) XN
 40   FORMAT(' ',10X,'ESTIMATED VALUE OF N =',F12.5)
      IF(XP(1).GT.0..AND.XP(2).GT.0.) GO TO 19
      IF(XP(1).GT.0.) D11 = 1.
      IF(XP(2).GT.0.) D22 = 1.
      IF(XP(1).GT.0..OR.XP(2).GT.0.) D12 = 0.
      DET = D11*D22 - D12**2
      TEMP = D11
      D11 = D22/DET
      D22 = TEMP/DET
      D12 = -D12/DET
      IF(XP(1).GT.0.) D11 = 0.
      IF(XP(2).GT.0.) D22 = 0.
      WRITE (6,45)
 45   FORMAT('0',10X,'ASYMPTOTIC COVARIANCE MATRIX OF ESTIMATES')
      WRITE (6,50)
 50   FORMAT('0',21X,'P',12X,'N')
      WRITE (6,55) D11, D12, D12, D22
 55   FORMAT(' ',10X,'P',2X,2E13.5/11X,'N',2X,2E13.5)
 19   CONTINUE
      IF(XP(2).GT.0.) GO TO 22
C
C     FIND INTEGER SOLUTION
C
      IF(XLK2.LE.XLK1) GO TO 21
      P1 = P2
      N1 = N2
      XLK1 = XLK2
 21   CONTINUE
      WRITE (6,70) P1, N1, XLK1
 70   FORMAT('0',10X,'SOLUTION FOR INTEGER N (PASCAL DISTRIBUTION):'/
     1 11X,'P =',F8.5,', N =',I5,', LOG-LIKELIHOOD =',F12.4)
 22   CONTINUE
C
C     COMPUTE GOODNESS OF FIT
C
      CALL GDFIT
      WRITE (6,65) XLK
 65   FORMAT('0',10X,'LOG-LIKELIHOOD =',F12.4)
      RETURN
 900  CONTINUE
      WRITE (6,905) XP(1), XP(2)
 905  FORMAT('0',10X,'INVALID PARAMETER SPECIFICATION'/18X,2F11.5)
      RETURN
 920  CONTINUE
      WRITE (6,925) V, XM1
 925  FORMAT('0',10X,'VARIANCE (',F10.5,') < MEAN (',F10.5,')'/
     1 11X,'NO FIT ATTEMPTED')
      RETURN
      END
```

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA-287 (Cin.)
(4-76)

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD. BEGIN SECTIONS HERE

```
C      SUBROUTINE TO FIT A TRUNCATED NEGATIVE BINOMIAL DISTRIBUTION
C         PROGRAMMERS: ANDERSON, HASSELBLAD              REVISED 10-20-80
C
       SUBROUTINE TNEGBN
       DIMENSION R(200)
C
C      NF IS AN ARRAY OF OBSERVED COUNTS
C      E IS AN ARRAY OF EXPECTED COUNTS
C      XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C      K IS THE VALUE CORRESPONDING TO THE LARGEST OBSERVED FREQUENCY
C      M IS THE LENGTH OF THE E ARRAY
C      INT IS AN OPTIONAL INPUT INTEGER.
C      NP IS THE NUMBER OF PARAMETERS ESTIMATED
C      IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C         AND IF = 2 FOR TRUNCATED DISTRIBUTIONS.
C      NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C      XN IS THE ESTIMATE OF THE PARAMETER N.
C      W IS THE ESTIMATE OF THE PARAMETER W, WHERE P = (1 - W)/W
C      P IS THE ESTIMATE OF THE PARAMETER P
C      SW IS THE FIRST PARTIAL WITH RESPECT TO W
C      SN IS THE FIRST PARTIAL WITH RESPECT TO N
C      R IS THE ARRAY OF CUMULATIVE FREQUENCIES
C
       COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
       WRITE (6,5) TITLE
 5     FORMAT('1',10X,20A4)
       WRITE (6,35)
 35    FORMAT('0',10X,'FIT TO A TRUNCATED NEGATIVE BINOMIAL')
       WRITE (6,15)
 15    FORMAT('0',10X,'P(X) = (N+X-1)!P**X/(X!(N-1)!(1+P)**(N+X)(1-1/(1+P
     1)**N))')
       WRITE (6,25) NTOT
 25    FORMAT('0',10X,'TOTAL SAMPLE SIZE =',I6)
       IF(XP(1).LT.0..OR.XP(2).LT.0.)GO TO 900
       IF = 2
       NT = 0
       KP = K + 1
       M = K + 2
       NP = 2
       IF(XP(1).GT.0.) NP = NP - 1
       IF(XP(2).GT.0.) NP = NP - 1
       XM = 0.
       DO 2 I = 2, KP
          CI = I - 1
          XM = XM + NF(I)*CI
          NT = NT + NF(I)
 2     CONTINUE
       TOT = NT
       XM = XM/TOT
       Q = 0.
       QS = 1. - XM
       DO 3 I = 2, KP
          Q = Q + NF(I)*QS**2
          QS = QS + 1.
 3     CONTINUE
       S = NF(2)/TOT
C
C      COMPUTE INITIAL ESTIMATES
C
       W = XM*(1. - S)*(TOT - 1.)/Q
       IF(W.LE.0.1) W = 0.1
       IF(W.GE.0.9) W = 0.9
       XN = (W*XM - S)/(1. - W)
       IF(XN.LE.0.5) XN = 0.5
       S = TOT*XM
       P = (1. -W)/W
       IF(XP(1).GT.0.) P = XP(1)
       IF(XP(2).GT.0.) XN = XP(2)
       IF(XP(1).GT.0.) W = 1./(1. + P)
C
```

BEGIN LAST LINE OF TEXT HERE
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin:)
(4-76)

```
C     FORM THE CUMULATIVE DISTRIBUTION OF OBSERVATIONS
C
      R(KP) = NF(KP)
      DO 4 I = 1, K
         J = KP - I
         R(J) = R(J+1) + NF(J)
 4    CONTINUE
      IF(XP(1).GT.0..AND.XP(2).GT.0.) GO TO 6
C
C     ITERATE USING A NEWTON-RAPHSON ITERATION SCHEME
C
      DO 8 IT = 1, 100
         D11 = 0.
         D22 = 0.
         Q = XN
         DO 9 I = 2, KP
            QS = R(I)/Q
            D11 = D11 + QS
            D22 = D22 + QS/Q
            Q = Q + 1.
 9       CONTINUE
         XK = TOT*XN
         WM = 1. - W
         Q = W**XN
         QS = 1. - Q
         AL = ALOG(W)
         SW = XK/(W*QS) - S/WM
         SN = TOT*AL/QS + D11
         IF(XP(1).GT.0.) SW = 0.
         IF(XP(2).GT.0.) SN = 0.
         QM = QS**2
C
C        COMPUTE TERMS OF INFORMATION MATRIX
C
         D12 = W*QM
         D11 = XK*(1. - (XN + 1.)*Q)/(W*D12) + S/WM**2
         D22 = D22 - TOT*AL**2*Q/QM
         D12 = TOT*(1. - (1. - XN*AL)*Q)/D12
         IF(XP(1).GT.0..OR.XP(2).GT.0.) D12 = 0.
         DET = D11*D22 - D12**2
         IF(DET.LE.0.) DET = D11*D22
         VW = D22/DET
         VN = D11/DET
         COV = D12/DET
C
C        CHECK FOR NEGATIVE VARIANCE ESTIMATES
C
         IF(VW.LT.0.) GO TO 7
         IF(VN.LT.0.) GO TO 7
C
C        ADJUST N AND W FOR NEXT ITERATION
C
         WP = W
         XNP = XN
         IF(XP(1).EQ.0.) W = W + SW*VW + SN*COV
         IF(XP(2).EQ.0.) XN = XN + SW*COV + SN*VN
         IF(W.GT.1.) W = (1. + WP)/2.
         IF(W.LE.0.) W = WP/2.
         IF(XN.LE.0.) XN = XNP/2.
C
C        CHECK FOR CONVERGENCE
C
         IF((ABS(SW) + ABS(SN)).LT.0.0001) GO TO 6
 8    CONTINUE
 7    CONTINUE
      WRITE (6,95)
 95   FORMAT('0',10X,'PROCEDURE DID NOT CONVERGE')
 6    CONTINUE
         IF(XP(1).EQ.0.) P = (1. - W)/W
C
```

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

```
C     COMPUTE LIKELIHOOD, EXPECTED COUNTS
C
      XLK = 0.
      ESUM = 0.
      XNC = XN
      Q = 1.+ P
      TERM = XN*ALOG(W) + ALOG(XN) + ALOG(1. - W) - ALOG(1. - W**XN)
      DO 12 I = 2, KP
         CI = I - 1
         E(I) = TOT*EXP(TERM)
         ESUM = ESUM + E(I)
         XLK = XLK + NF(I)*TERM
         XNC = XNC + 1.
         TERM = TERM + ALOG(1. - W) + ALOG(XNC) - ALOG(CI + 1.)
 12   CONTINUE
      E(M) = TOT - ESUM
C
C     WRITE OUT ESTIMATES AND THEIR VARIANCES
C
      IF(XP(1).GT.0.) WRITE (6,60) P
 60   FORMAT('0',10X,'SPECIFIED VALUE OF P =',F12.5)
      IF(XP(1).EQ.0.) WRITE (6,65) P
 65   FORMAT('0',10X,'ESTIMATED VALUE OF P =',F12.5)
      IF(XP(2).GT.0.) WRITE (6,70) XN
 70   FORMAT(' ',10X,'SPECIFIED VALUE OF N =',F12.5)
      IF(XP(2).EQ.0.) WRITE (6,75) XN
 75   FORMAT(' ',10X,'ESTIMATED VALUE OF N =',F12.5)
      IF(XP(1).GT.0..AND.XP(2).GT.0.) GO TO 13
      D11 = D11*W**4 + 2.*SW*W**3
      D12 = -D12*W**2
      IF(XP(1).GT.0.) D11 = 1.
      IF(XP(2).GT.0.) D22 = 1.
      IF(XP(1).GT.0..OR.XP(2).GT.0.) D12 = 0.
      DET = D11*D22 - D12**2
      TEMP = D11
      D11 = D22/DET
      D22 = TEMP/DET
      D12 = D12/DET
      IF(XP(1).GT.0.) D11 = 0.
      IF(XP(2).GT.0.) D22 = 0.
      WRITE (6,80)
 80   FORMAT('0',10X,'ASYMPTOTIC COVARIANCE MATRIX OF ESTIMATES')
      WRITE (6,85)
 85   FORMAT('0',21X,'P',12X,'N')
      WRITE (6,90) D11, D12, D12, D22
 90   FORMAT(' ',10X,'P',2X,2E13.5/11X,'N',2X,2E13.5)
 13   CONTINUE
C
C     COMPUTE GOODNESS OF FIT
C
      CALL GDFIT
      WRITE (6,40) XLK
 40   FORMAT('0',10X,'LOG-LIKELIHOOD =',F12.4)
      RETURN
 900  CONTINUE
      WRITE (6,905) XP(1), XP(2)
 905  FORMAT('0',10X,'INVALID PARAMETER SPECIFICATION'/18X,2F11.5)
      RETURN
      END
```

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin.) (4-76)

```
C     SUBROUTINE TO FIT A LOGARITHMIC DISTRIBUTION     REVISED  7-30-80
C        PROGRAMMERS: ANDERSON, HASSELBLAD
      SUBROUTINE LOGRTH
      COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
C
C     NF IS AN ARRAY OF OBSERVED COUNTS
C     E IS AN ARRAY OF EXPECTED COUNTS
C     XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C     K IS THE VALUE CORRESPONDING TO THE LARGEST OBSERVED FREQUENCY
C     M IS THE LENGTH OF THE E ARRAY
C     INT IS AN OPTIONAL INPUT INTEGER.
C     NP IS THE NUMBER OF PARAMETERS ESTIMATED
C     IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C        AND IF = 2 FOR TRUNCATED AND LOGARITHMIC DISTRIBUTIONS.
C     NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C
      WRITE (6,30) TITLE
 30   FORMAT('1',10X,20A4)
      WRITE (6,5)
 5    FORMAT('0',10X,'FIT TO A LOGARITHMIC DISTRIBUTION')
      KP = K + 1
      NP = 1
      IF(XP(1).GT.0)NP = 0
      M = K + 2
      IF = 2
      NT = 0
      NUM = 0
C
C     COMPUTE SAMPLE SIZE AND MEAN
C
      XM1 = 0.
      DO 2 I = 2, KP
         CI = I - 1
         NT = NT + NF(I)
         XM1 = XM1 + NF(I)*CI
 2    CONTINUE
      TOT = NT
      XM1 = XM1/TOT
      WRITE (6,10) NT
 10   FORMAT('0',10X,'TOTAL SAMPLE SIZE =',I6)
C
C     CHECK FOR IMPROPER PARAMETER VALUE
C
      THETA = XP(1)
      IF(XP(1).GT.0.) GO TO 3
      TEMP = ALOG(XM1)
C
C     COMPUTE INITIAL ESTIMATE
C
      THETA = 1. - 1./(1. + ((5./3. - TEMP/16.)*(XM1 - 1.) + 2.)*TEMP)
C
C     USE A NEWTON-RAPHSON ITERATION SCHEME
C
      DO 4 I = 1, 50
         TEMP = 1. - THETA
         THETP = THETA - (XM1 + THETA/(TEMP*ALOG(TEMP)))*(((TEMP*
     1    ALOG(TEMP))**2/(ALOG(TEMP) + THETA)))
         IF(ABS(THETP-THETA).LE.0.00001) GO TO 6
         THETA = THETP
 4    CONTINUE
      WRITE (6,95)
 95   FORMAT('0',10X,'PROCEDURE DID NOT CONVERGE')
 6    CONTINUE
      THETA = THETP
 3    CONTINUE
      IF(XP(1).EQ.0.) WRITE (6,15) THETA
 15   FORMAT('0',10X,'ESTIMATED VALUE OF THETA =',F8.4)
      IF(XP(1).GT.0.) WRITE (6,25) THETA
 25   FORMAT('0',10X,'SPECIFIED VALUE OF THETA =',F8.4)
C
```

```
C       COMPUTE ESTIMATED VARIANCE OF ESTIMATE
C
        ALT =-ALOG(1. - THETA)
        TEMP = 1. - THETA
        VAR = THETA*TEMP**2*ALOG(TEMP)/(ALT*(ALOG(TEMP) + THETA))
        IF(XP(1).EQ.0) WRITE (6,45) VAR
 45     FORMAT(' ',10X,'ASYMP. VAR. OF THETA =',E13.5)
        CUM = 0.
C
C       COMPUTE EXPECTED COUNTS, LOG-LIKELIHOOD
C
        E(1) = 0.
        XL = 0.
        DO 7 I = 2, KP
           CI = I-1
           TERM = THETA**(I - 1)/(CI*ALT)
           E(I) = TERM*TOT
           XL = XL + NF(I)*ALOG(TERM)
           CUM = CUM + E(I)
 7      CONTINUE
        E(M) = TOT - CUM
C
C       COMPUTE GOODNESS OF FIT
C
        CALL GDFIT
        WRITE (6,20) XL
 20     FORMAT('0',10X,'LOG-LIKELIHOOD  =',F12.4)
        RETURN
 900    CONTINUE
        WRITE (6,905) XP(1)
 905    FORMAT('0',10X,'INVALID PARAMETER SPECIFICATION'/18X,F11.5)
        RETURN
        END
```

```
C      SUBROUTINE TO COMPUTE A CHI-SQUARE GOODNESS OF FIT TEST
C         PROGRAMMER: HASSELBLAD                          REVISED 10-20-80
C
       SUBROUTINE GDFIT
       COMMON TITLE(20),NF(200),E(200),XP(4),K,M,INT,NP,IF,NTOT
C      THIS ROUTINE CONMPUTES A CHI-SQUARE GOODNESS OF FIT TEST BASED ON
C      THE OBSERVED AND EXPECTED FREQUENCIES.  THE CELLS ARE COLLAPSED
C      FROM THE RIGHT AND LEFT TO GET A MINIMUM EXPECTED FREQUENCY OF
C      EXMIN.
C
C      NF IS AN ARRAY OF OBSERVED COUNTS
C      E IS AN ARRAY OF EXPECTED COUNTS
C      XP IS AN ARRAY OF OPTIONAL INPUT CONSTANTS
C      M IS THE LENGTH OF THE E ARRAY
C      INT IS AN OPTIONAL INPUT INTEGER.
C      NP IS THE NUMBER OF PARAMETERS ESTIMATED
C      IF IS THE INITIAL FREQUENCY USED AND IS 1 FOR MOST DISTRIBUTIONS;
C         AND IF = 2 FOR TRUNCATED DISTRIBUTIONS.
C      NTOT IS THE SUM OF THE COUNTS IN THE NF ARRAY
C
C
       DATA EXMIN /5./
C
C      EXMIN IS THE MINIMUM EXPECTED FREQUENCY ALLOWED AFTER COLLAPSING.
C
       NDF = 1 - NP
       KP = K + 1
       TOT = 0.
       DO 9 I = IF, M
          TOT = TOT + E(I)
 9     CONTINUE
       ESUM = 0.
       NSUM = 0.
       WRITE (6,5)
 5     FORMAT('0',13X,'CHI-SQUARE GOODNESS OF FIT TEST'/'0',10X,
      1 'VALUE  OBSERVED  EXPECTED  CHI-SQUARE'/' ')
C
C      SUM INITIAL EXPECTED FREQUENCIES UNTIL THEY REACH EXMIN
C
       DO 2 I = IF, M
          J = I
          IT = I - 1
          ESUM = ESUM + E(I)
          NSUM = NSUM + NF(I)
          IF(ESUM.GE.EXMIN) GO TO 3
          WRITE (6,10) IT, NF(I), E(I)
10     FORMAT(' ',10X,I3,I9,F12.2,7X,'--')
 2     CONTINUE
       GO TO 900
 3     CONTINUE
       CHIS = (NSUM - ESUM)**2/ESUM
       TCHS = CHIS
       WRITE (6,15) IT, NF(J), E(J), CHIS
15     FORMAT(' ',10X,I3,I9,F12.2,F10.2)
       L = J + 1
C
C      WRITE OUT INTERMEDIATE COUNTS UNTIL REMAINING EXPECTED FREQUENCIES
C      ARE LESS THAN XMIN.
C
       DO 4 I = L, M
          J = I
          IT = I - 1
          ESUM = ESUM + E(I)
          NSUM = NSUM + NF(I)
          IF(TOT-ESUM.LT.EXMIN) GO TO 6
          NDF = NDF + 1
          CHIS = (NF(I) - E(I))**2/E(I)
          TCHS = TCHS + CHIS
          WRITE (6,15) IT, NF(J), E(J), CHIS
 4     CONTINUE
```

```
    6   CONTINUE
        ESUM = TOT - ESUM + E(J)
        CHIS = (NTOT - NSUM + NF(J) - ESUM)**2/ESUM
        TCHS = TCHS + CHIS
        IF(J.LT.M.OR.KP.EQ.M) WRITE(6,15) IT, NF(J), E(J), CHIS
        IF(J.EQ.M.AND.KP.LT.M) WRITE (6,25) IT, NF(J), E(J), CHIS
   25   FORMAT(' ',10X,I3,'+',I8,F12.2,F10.2)
        IF(J.GE.M) GO TO 7
        L = J + 1
C
C       WRITE OUT REMAINING EXPECTED FREQUENCIES AND COUNTS
C
        DO 8 I = L, M
           J = I
           IT = I - 1
           IF(J.LT.M.OR.KP.EQ.M) WRITE (6,10) IT, NF(I), E(I)
           IF(J.EQ.M.AND.KP.LT.M) WRITE (6,30) IT, NF(I), E(I)
   30      FORMAT(' ',10X,I3,'+',I8,F12.2,7X,'--')
    8   CONTINUE
    7   CONTINUE
        PVAL = CHISQ(TCHS,NDF)
        WRITE (6,20) TCHS, NDF, PVAL
   20   FORMAT('0',10X,'CHI-SQUARE =',F8.2,', D.F. =',I3,', P =',F6.4)
        RETURN
  900   CONTINUE
        WRITE (6,905)
  905   FORMAT('0',10X,'INSUFFICIENT EXPECTED FREQUENCIES')
        RETURN
        END
```